

AthaMap, eine Datenbank für die  
Identifizierung kombinatorischer Elemente  
durch Colokalisationsanalysen von Transkriptionsfaktorbindungsstellen

Von der Fakultät für Lebenswissenschaften  
der Technischen Universität Carolo-Wilhelmina  
zu Braunschweig  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
(Dr. rer. nat.)  
genehmigte  
Dissertation

Von Nils Ole Steffens  
aus Hamburg

1. Referent

apl. Professor Dr. Reinhard Hehl

2. Referent

Professor Dr. Edgar Wingender

eingereicht am:

12.06.2006

mündliche Prüfung (Disputation) am:

16.02.2007

Druckjahr 2007

## **Vorveröffentlichungen der Dissertation**

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor, in folgenden Beiträgen vorab veröffentlicht:

## **Publikationen**

Hehl, R., Steffens N.O., Wingender E. (2004) Isolation and analysis of gene regulatory sequences. In: Handbook of Plant Biotechnology, Klee H, and Christou P, eds., (Wiley and Sons Ltd.), pp. 81-102.

Steffens N.O., Galuschka C., Schindler M., Bülow L., Hehl R. (2004) AthaMap: an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome.

Nucleic Acids Res. (32): D368-72.

Steffens N.O., Galuschka C., Schindler M., Bülow L., Hehl R. (2005) AthaMap web tools for database-assisted identification of combinatorial *cis*-regulatory elements and the display of highly conserved transcription factor binding sites in *Arabidopsis thaliana*.

Nucleic Acids Res. (33): W397-402.

Bülow L., Steffens N. O., Galuschka C., Schindler M. and Hehl R. (2006) AthaMap: from *in silico* data to real transcription factor binding sites.

In Silico Biol. 6, 0023

## **Tagungsbeiträge:**

Steffens N.O., Hehl, R. Colokalisierung von Transkriptionsfaktor-Bindungsstellen im *Arabidopsis* Genom gibt Hinweise auf potentielle TF-Dimerisierung. (Poster) 16

Tagung: „Molekularbiologie der Pflanzen“, Dabringhausen 2003, Deutschland

Steffens N.O., Galuschka C., Bülow L., Schindler M., Hehl R. AthaMap, an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. "German Conference on Bioinformatics", München 2003, Deutschland

Steffens N.O., Galuschka C., Bülow L., Schindler M., Hehl R. AthaMap, an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. (Update) "The *Arabidopsis* conference", Berlin 2004, Deutschland

Galuschka C., Steffens N.O., Bülow L., Hehl R. Non-Random Distribution of Transcription Factor Binding Sites in the *Arabidopsis thaliana* Genome " The *Arabidopsis* conference", Berlin 2004, Deutschland

Steffens N.O., Galuschka C., Bülow L., Schindler M., Hehl R. AthaMap, an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. (Update) "Botanikertagung 2004", Braunschweig 2004, Deutschland

Galuschka C., Steffens N.O., Bülow L., Hehl R. Non-Random Distribution of Transcription Factor Binding Sites in the *Arabidopsis thaliana* Genome (Update) "Botanikertagung 2004", Braunschweig 2004, Deutschland

Hehl R., Bülow L., Cerff R., Galuschka C., Grohmann J., Hänsch R., Lisson R., Mendel R.R., Palme K., Sell S., Steffens N.O. Forward and reverse genetic approaches to study CO<sub>2</sub> dependent anaerobic signal transduction " 8th Conference of the International Society for Plant Anaerobiosis", Perth 2004, Australien

Steffens N.O., Galuschka C., Schindler M., Bülow L., and Hehl R. The AthaMap database resources for identification of *cis*-regulatory elements and the prediction of coregulated genes in *Arabidopsis thaliana*. "Plant Biology 2005", Seattle 2005, USA

## Verzeichnis der Abkürzungen

<b>Abkürzung</b>	<b>Bedeutung</b>
Abb.	Abbildung
ABA	Abscisinic acid
ABRE	Abscisinic acid response element
AGI	Arabidopsis Genome Initiative
bp	Basenpaare
CBF	CCAAT-Box bindender Faktor
CE	Coupling element
DNA	Desoxyribonucleic acid
HTML	Hypertext markup language
INR	Initiatorelement zur Initiation der Transkription
kb	Kilobasen
MBD	Methylated DNA-binding domain
PC	Personal Computer
RBSS	Random binding site selection assay
SQL	Structured query language
TAIR	The Arabidopsis information resource ( <a href="http://www.arabidopsis.org">www.arabidopsis.org</a> )
TIC	Transkriptions-Initiationskomplex
TIGR	The institute for genomics research ( <a href="http://www.tigr.org">www.tigr.org</a> )
TBP	TATA-box bindendes Protein
TF	Transkriptionsfaktor
TFBS	Transkriptionsfaktorbindungsstelle(n)
TSS	Transkriptionsstartpunkt
UTR	Untranslatierte Region

# Inhaltsverzeichnis

<b>1.</b>	<b>Einleitung</b>	<b>1</b>
1.1	Transkriptionsfaktoren (TFs) regulieren die Gentranskription	1
1.2	Epigenetische Regulation der Gentranskription	4
1.3	Allgemeine TFs der Eukaryonten: TBP und CBF	5
1.3.1	Das TATA-Box bindende Protein	5
1.3.2	Der CCAAT-Box bindende Faktor	7
1.4	Transkriptionsfaktorbindungsstellen können durch Matrizen repräsentiert werden	8
1.5	Die Bindungsstellen interagierender TFs können kombinatorische Elemente bilden	9
1.6	Colokalisationsanalyse von TFBS kann Hinweise auf potenzielle kombinatorische Elemente geben	10
1.7	Motivation zur Erstellung von AthaMap, einer genomweiten Datenbank <i>cis</i> -regulatorischer Elemente in <i>Arabidopsis thaliana</i>	12
1.8	Weitere Datenbanken für die Analyse der pflanzlichen Genregulation und <i>cis</i> -regulatorischer Elementen	13
1.9	Aufbau Abscisinsäure-responsiver Elemente (ABREs)	15
<b>2.</b>	<b>Material und Methoden</b>	<b>18</b>
2.1	Genom-Daten	18
2.2	Material für Matrizenscreenings	18
2.3	Identifizierung potentieller TFBS	19
2.4	Import von TFBS in die AthaMap Datenbank	22
2.5	Identifizierung von kombinatorischen Elementen	25
2.6	Import von kombinatorischen Elementen	27
2.7	Allgemeiner Aufbau der AthaMap Datenbank	28
2.8	Bestimmung der Position von TFBS in AthaMap relativ zum ATG	30
2.9	Das Tool CECleaner zur Entfernung von überlappenden kombinatorischen Elementen in AthaMap	31
2.10	Bestimmung der relativen Position zur Genstruktur von Einzel- und kombinatorischen Elementen	33

2.11	Positionell eingeschränkte Annotation von TBP und CBF-Bindungsstellen in AthaMap2004 .....	35
2.12	Entfernung positionsgleicher TBP-Bindungsstellen .....	36
2.13	Bestimmung der durchschnittlichen Länge der 5'UTRs inkl. Introns .....	37
2.14	Selektion zufällig gewählter Gene mittels RandomGeneSelector .....	38
2.15	Korrelation von <i>in silico</i> Daten mit Microarray-basierten Genexpressionsprofilen .....	40
2.16	Implementierung des AthaMap Webinterfaces .....	41
2.17	Statistische Auswertung mittels t-Test .....	41
2.18	Test auf Normalverteilung mithilfe des Kolmogorov-Smirnov-Testes .....	42
<b>3.</b>	<b>Ergebnisse</b> .....	<b>43</b>
3.1	Die AthaMap Datenbank .....	43
3.1.1	Dateninhalt von AthaMap .....	43
3.1.2	Navigation in AthaMap .....	46
3.1.3	Visualisierung der <i>Arabidopsis</i> -Genomsequenz in AthaMap .....	47
3.1.4	Webtools in AthaMap: Restriktion von TFBS auf stark konservierte Bindungsstellen und Colokalisationsanalyse .....	48
3.2	Identifizierung putativ funktioneller TATA-Boxen .....	51
3.2.1	Das Genom von <i>Arabidopsis thaliana</i> enthält mehr TATA-Boxen als erwartet .....	51
3.2.2	TATA-Boxen kumulieren in Upstreambereichen annotierter Gene .....	51
3.2.3	TATA-Boxen halten einen bevorzugten Abstand von 35bp $\pm$ 5bp zum TSS ein .....	52
3.2.4	Kriterien für die Identifizierung und Annotation von TATA-Boxen im Upstreambereich von Genen ohne bekannten TSS .....	54
3.2.5	Die Mehrzahl der Gene in <i>Arabidopsis</i> besitzt keine TATA-Box .....	56
3.2.6	Bei Genen mit unbekanntem TSS treten häufiger mehrfache TATA-Boxen auf .....	56
3.2.7	CCAAT-Boxen besitzen keinen stark konservierten Abstand zum ATG oder TSS .....	58
3.3	Prädiktion von potenziell ABA-regulierten Genen anhand von Colokalisationsanalysen .....	61

3.3.1	Ein Großteil der Gene mit ABF1-TBP-Colokalisationen ist ABA-induzierbar	61
3.3.2	Die ABF1-Bindungsstelle von ABF1-TBP-Colokalisationen tritt bevorzugt im upstream-Bereich auf	66
3.3.3	Bei intergenischen ABF1-TBP-Colokalisationen tritt die ABF1-Bindungsstelle bevorzugt distal zur TBP-Bindungsstelle auf	68
3.3.4	ABF1-TBP-Colokalisation definiert ein neues ABRE	72
3.3.5	Klassische ABREs können durch ABF1-Opaque2-Colokalisationen identifiziert werden	73
3.3.6	Das Auftreten einer einzelnen ABF1-Bindungsstelle ohne CE oder TATA-Box reicht nicht aus, um die ABA-Induzierbarkeit von Genen zu vermitteln	74
3.3.7	Die Gene der Prädiktionsgruppe sind auch durch osmotischen Streß und Salzstreß induzierbar	76
3.3.8	Die Induktionsbedingungen von ABF3 korrelieren mit den Expressionsbedingungen der Gene der Prädiktionsgruppe	78
<b>4.</b>	<b>Diskussion</b>	<b>79</b>
4.1	AthaMap ist die erste genomweite Karte von TFBS in <i>Arabidopsis thaliana</i>	79
4.2	AthaMap enthält putativ funktionelle TATA-Boxen	81
4.2.1	Die Lokalisierung von ATG- und TSS-annotierten TATA-Boxen	83
4.2.2	Multiple TATA-Boxen treten häufiger vor Genen ohne bekannten TSS auf	84
4.2.3	Die Mehrzahl der <i>Arabidopsis</i> -Gene besitzt keine TATA-Box im Promotor	87
4.3	Colokalisationsanalysen von ABF1-Bindungsstellen mit TATA- und G-Boxen geben Hinweise auf ein neuartiges ABRE	88
4.3.1	Die Lokalisierung der ABF1-Bindungsstelle ist bedeutsam für die Funktionalität des neuartigen ABREs	89
4.3.2	Bei intergenischen ABF1-TBP-Colokalisationen lokalisiert die ABF1-Bindungsstelle bevorzugt distal	90
4.3.3	Das ABF1/TBP ABRE wird wahrscheinlich von ABF3 reguliert	90



<b>5.</b>	<b>Zusammenfassung.....</b>	<b>92</b>
<b>6.</b>	<b>Literaturverzeichnis.....</b>	<b>94</b>
<b>7.</b>	<b>Anhang.....</b>	<b>110</b>
7.1	Herleitung der verwendeten Matrizen.....	110
7.2	Kontrollgruppen zufällig ausgewählter Gene zum Vergleich mit der ABF1-TBP-Prädiktionsgruppe.....	131
7.3	ABF1-Downstreamgruppe und Vergleichsgruppen zufällig ausgewählter Gene.....	136
7.4	Gene der ABF1_O2_5_50bp Prädiktionsgruppe.....	142
7.5	Gene der ABF1_O2_51_100bp Prädiktionsgruppe.....	143
7.6	Gene der ABF1_O2_101_150bp Prädiktionsgruppe.....	144
7.7	Gene der ABF1_O2_151_200bp Prädiktionsgruppe.....	145
7.8	Gene der ABF1_O2_201-250bp Prädiktionsgruppe.....	146
7.9	Kontrollgruppen mit angepasster Gruppengröße für Gene mit ABF1-TBP-Colokalisation aber ohne klassisches G-Box ABRE.....	147
7.9.1	t-Test auf Signifikanz nach Subtraktion der Gene mit klassischem ABRE.....	152
7.10	ABF1_TSS_0_50bp-Gruppe und Vergleichsgruppen zufällig ausgewählter Gene.....	153

## 1. Einleitung

### 1.1 Transkriptionsfaktoren regulieren die Gentranskription

Pflanzen sind als sessile Organismen noch stärker als autonom mobile Organismen darauf angewiesen, sich den wechselnden Anforderungen ihrer Umgebung anzupassen. Die Adaption an Umweltbedingungen kann physiologischer, metabolischer oder morphologischer Natur sein und erfordert als ersten Schritt die Wahrnehmung der Umgebungsbedingungen. Daran anschließend wird eine Signaltransduktionskette in Gang gesetzt, die letztendlich zu einer Anpassung der Genexpression als Reaktion auf die sich verändernde Umwelt führt. Beispielsweise können Pflanzen auf Wassermangel durch verstärkte Expression von Wassertransportproteinen reagieren [Jang et al., 2004], ihre Genexpression der veränderlichen Lichtsituation anpassen [Jiao et al., 2003], den Befall durch Pathogene wahrnehmen und bekämpfen [Dowd et al., 2004] und so auf Herausforderungen ihrer Umgebung reagieren.

Auf der anderen Seite erfordert der Entwicklungszyklus der Pflanze eine fein abgestimmte Expression der Gene in den verschiedenen Pflanzengeweben bzw. zu unterschiedlichen Zeitpunkten [Obayashi et al., 2004]. Die Identität der verschiedenen Pflanzenorgane wird durch organspezifische Genexpression festgelegt [Obayashi et al., 2004], gleichermaßen benötigen die jeweiligen Organe bestimmte charakteristische Proteine, um ihre Aufgaben erfüllen zu können.

Sowohl Anpassung an die Umwelt als auch Entwicklung erfordern eine dynamische Regulation der Genexpression, deren erster Schritt die Gentranskription ist.

Die DNA von Eukaryonten liegt normalerweise in Nukleosomen kompaktiert vor, was zu einer generellen Unterdrückung der Transkription führt [Hsieh & Fischer, 2004].

Um die Transkription zu aktivieren, muss das Chromatin remodeliert werden und der Transkriptions-Initiationskomplex zum Transkriptionstartpunkt eines Gens rekrutiert werden. Die Prozesse, die zur Remodelierung von Heterochromatin zu Euchromatin führen, sind unter Abschnitt 1.2. („Epigenetische Regulation“) näher beschrieben.

Transkriptionsfaktoren erfüllen unterschiedliche Aufgaben sowohl bei der Neuorganisation des Chromatins als auch bei der Rekrutierung des Transkriptions-Initiationskomplexes.

Der erste Schritt bei der Regulation der Transkriptionsaktivität eines Gens ist die Bindung von Transkriptionsfaktoren (TFs) an spezifische Sequenzmotive der DNA im regulatorischen Bereich des Gens. Der regulatorische Bereich, in dem die Bindung von TFs erfolgt, befindet sich meist im nicht-transkribierten 5'-Bereich vor dem Transkriptionsstart des Gens, es sind allerdings auch regulatorische Elemente in Introns bekannt [Hong et al., 2003].

Die Bindung eines TFs an die DNA kann aktivierende oder reprimierende Wirkung auf die Transkription des Gens ausüben [Fujita et al., 2005; Uehara et al., 2005; Singh, 1998]; die entsprechenden TFs werden als Transkriptionsaktivatoren oder –repressoren bezeichnet. Regulatorische Proteine, die ebenfalls die Transkription aktivieren oder reprimieren können, ohne direkt an die DNA zu binden, werden als Coaktivatoren bzw. Corepressoren bezeichnet. Diese Cofaktoren interagieren mit DNA-bindenden Proteinen bzw. modifizieren DNA-bindende Proteine und ermöglichen so die Bildung des DNA-bindenden Komplexes [Pan et al., 1999]. Eine dritte Klasse von TFs sind die allgemeinen Transkriptionsfaktoren, die Bestandteile des Polymerase-II–Transkriptions-Initiationskomplexes sind, der für die Transkription der meisten proteincodierenden Gene benötigt wird.

Bei den Aktivatoren und Repressoren erfolgt die Interaktion mit der DNA über ihre DNA-Bindungsdomäne, einen diskreten Bereich des Proteins, der die Erkennung einer bestimmten DNA-Sequenz ermöglicht und die spezifische Bindung an diese Sequenz vermittelt. Anhand der Homologie ihrer DNA-Bindungsdomäne werden TFs in verschiedene Klassen eingeteilt [Riechmann et al., 2000]. Die TFs einer Klasse besitzen meist eine ähnliche Bindungsspezifität, da ihre DNA-Bindungsdomänen ähnlich sind. Einige TFs besitzen mehr als eine DNA bindende Domäne, beispielsweise die TFs MYB.PH3 oder RAV1. In beiden Fällen liegen zwei DNA-Bindungsdomänen vor, wobei im Falle von MYB.PH3 nur eine von beiden Bindungsdomänen an die DNA gebunden vorliegt [Solano et al., 1995], während die beiden Bindungsdomänen von RAV1 gleichzeitig Protein-DNA-Kontakte ausbilden [Kagaya et al. 1999].

Ein weiteres Kennzeichen von TFs ist ihre Effektdomäne, die die Transkription des regulierten Gens entweder stimuliert oder reprimiert.

Neben einer oder mehrerer DNA-Bindungsdomänen kann ein TF auch eine oder mehrere Interaktionsdomänen besitzen, welche die Protein-Protein-Interaktion mit anderen TFs oder regulatorischen Proteinen ermöglichen [Singh 1998]. Einige TFs

benötigen eine solche Interaktion, um überhaupt zur DNA-Bindung in der Lage zu sein. So sind beispielsweise die bZIP-Faktoren GBF oder CPRF nur jeweils als Homodimer in der Lage, an die DNA zu binden [Jakoby et al., 2002]. Andere TFs können als Monomer an die DNA binden, entfalten ihre spezifische regulatorische Wirkung aber nur im Falle einer Interaktion mit anderen räumlich eng benachbarten Proteinen [Diaz et al., 2005]. Die Spezifität der Genregulation wird also nicht nur durch die spezifische Erkennung der DNA-Bindungssequenz vermittelt, sondern zumindest teilweise auch durch Protein-Protein-Interaktionen.

Darüber hinaus gibt es ein breites Spektrum weiterer Mechanismen, die die Aktivität von TFs regulieren können. Die Expression vieler TFs wird zeitlich und/oder räumlich stringent reguliert [Nagpal et al., 2005], so dass ein TF beispielsweise nur in bestimmten Geweben oder zu bestimmten Zeitpunkten für kombinatorische Interaktionen zur Verfügung steht und folglich die Zielgene des TFs derselben zeitlichen oder räumlichen Regulation folgen. Ein anderer häufiger biochemischer Regulationsmechanismus ist die Phosphorylierung von TFs [Furihata et al., 2006]. Weitere, etwas speziellere Regulationsmechanismen sind z.B. Redoxregulation [Heine et al., 2004] und die Bindung von Liganden [Despres et al., 2000], die die Aktivität eines TFs regulieren können.

Die ca. 1500 TFs von *Arabidopsis* lassen sich in mindestens 13 TF-Hauptfamilien einteilen [Riechmann et al., 2000]. Die größten TF-Familien in *Arabidopsis* sind die AP2/EREBP-, MYB(R1)R2R3- und die bHLH-Familie.

Einige TF-Familien sind pflanzenspezifisch, wie die AP2/EREBP-Familie, die WRKY-, Trihelix-, ARF- oder die NAC-Familie. Andere TF-Familien sind in allen Organismenreichen zu finden, wie z.B. die bZIP-Familie. Etwa 45% der TFs in *Arabidopsis* sind pflanzenspezifische TFs.

Mit ca. 1500 TF-codierenden Genen machen Transkriptionsfaktoren ca. 5,9% aller *Arabidopsis*-Gene aus. Damit ist der Anteil der TF-codierenden Gene im *Arabidopsis*-Genom größer als beispielsweise im Genom von *Caenorhabditis elegans* oder *Drosophila melanogaster*. Ein bedeutender Anteil der *Arabidopsis*-TFs ist wahrscheinlich durch Genduplikation entstanden. Abhängig von der TF-Familie besitzen durchschnittlich ca. 40% der TFs in *Arabidopsis* ein eng verwandtes homologes Gen [Riechmann et al., 2000].

## 1.2 Epigenetische Regulation der Gentranskription

Die genomische DNA von Eukaryonten liegt nicht frei im Zellkern vor, sondern in kompakter Form, den Nukleosomen. Nukleosomen sind Protein-DNA-Komplexe, deren Protein-Anteil von den vier Histonen H2A, H2B, H3 und H4 gebildet wird. Jeweils zweimal vier Histon-Monomere bilden ein Octamer, auf das der DNA-Strang aufgewickelt vorliegt (Abb. 1).

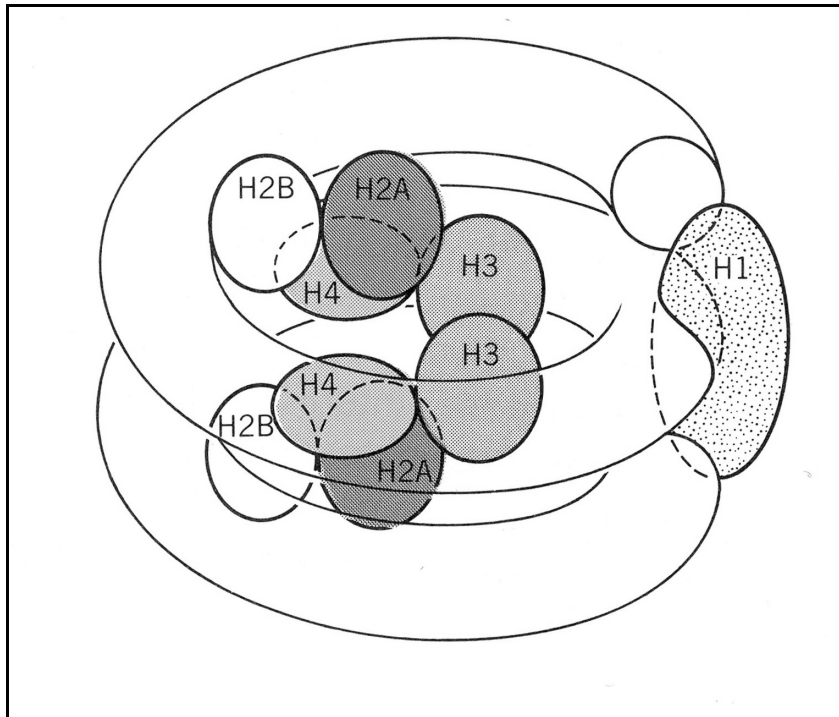


Abb. 1: Histone und DNA bilden Nukleosomen (aus: Lewin; Genes IV)

Da die DNA in dieser Form unzugänglich für Prozesse wie Transkription, Replikation oder Rekombination ist, muss der Nukleosom-Komplex zunächst teilweise aufgelöst werden. Dieser Prozess wird Remodelierung genannt und bedeutet die Umwandlung von stark kondensiertem Heterochromatin in decondensiertes Euchromatin, das für die Transkription und andere Prozesse zur Verfügung steht. An der Remodelierung sind verschiedene biochemische Prozesse beteiligt, die zum Teil synergistisch wirken können bzw. miteinander interagieren.

Ein wesentlicher Regulationsmechanismus für die Remodelierung ist die posttranslationale Modifizierung der Histone durch Acetylierung bzw. Deacetylierung. Die Acetylierung neutralisiert die positive Ladung der Histone und schwächt daher

die Protein-DNA-Wechselwirkungen, so dass die Ablösung der DNA von den Histonen erleichtert wird [Fransz & de Jong, 2002].

Ein weiterer posttranslationaler Regulationsmechanismus ist die Methylierung der Histone an ihren Lysinresten. Anders als bei der Acetylierung entscheidet hier nicht nur das Vorhandensein von Methylierung, sondern das Methylierungsmuster der Histone, ob der entsprechende DNA-Bereich aktiv oder inaktiv ist [Jasencakova et al., 2003].

Epigenetische Regulation findet jedoch auch auf DNA-Ebene statt; die Methylierung der DNA an Cytosin und Guanin-Resten ist eine der am häufigsten auftretenden epigenetischen Modifikationen [Finnegan et al., 2000]. DNA-Methylierung führt im Allgemeinen zum Erliegen der Transkriptionsaktivität („Silencen“) in dem betroffenen Bereich und kann bei der Zellteilung an die Tochterzellen weitervererbt werden.

In tierischen Systemen kann methylierte DNA von MBD-Proteinen (Methylated DNA binding domain) erkannt werden, welche dann Histondeacetylasen rekrutieren, die wiederum die Histone modifizieren und so eine synergistische Interaktion beider epigenetischer Regulationsmechanismen bewirken. In *Arabidopsis* wurde bisher nur die Existenz von MBD-homologen Proteinen nachgewiesen, die ähnliche Aufgaben erfüllen [Scebba et al., 2003; Zemach & Grafi 2003].

Parallel dazu gibt es in Pflanzen eine Interaktion von DNA-Methylierung und Histon-Methylierung; die Mutation der Histonmethyltransferase KYP führt dazu, dass die Gene CNG und PAI2, die normalerweise methyliert vorliegen, demethyliert werden [Jackson et al., 2002; Malagnac et al., 2002].

### **1.3 Allgemeine TFs der Eukaryonten: TBP und CBF**

#### **1.3.1 Das TATA-Box bindende Protein**

Proteincodierende Gene werden normalerweise von der RNA-Polymerase II transkribiert und besitzen einen „Corepromotor“, der die für die Erkennung als Promotor nötigen Elemente enthält. Am Corepromotor wird der Transkriptions-Initiationskomplex (TIC) assembliert und gebunden. Der TIC besteht aus allgemeinen TFs, assoziierten Faktoren und der RNA-Polymerase II mit einer ungefähren

## 1. Einleitung

Gesamtmasse von 2.500.000 Da. Die allgemeinen TFs TFIIA, TFIIB, TFIID, TFIIE, TFIIF und TFIIH sind zusammen mit zahlreichen weiteren Cofaktoren Bestandteile dieses Komplexes (Abb. 2) [Martinez 2002].

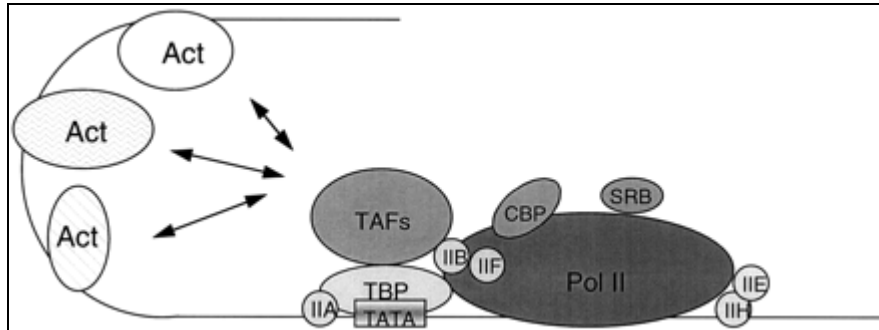


Abb. 2: Aufbau des Transkriptions-Initiationskomplexes nach Singh [1998]. Abkürzungen: Act: Transkriptionsaktivatoren. TAF: TATA-box associated factor. IIA, IIB, IIF, IIE, IIH: assoziierte TFs des Prä-Initiationskomplexes. Pol II: RNA-Polymerase II. SRB, CBP: Zusätzliche Faktoren

An dem Aufbau eines stabilen Prä-Initiationskomplexes im Bereich des Corepromotors sind gleichermaßen DNA-Protein-Interaktionen wie auch Protein-Protein-Interaktionen beteiligt [Orphanides et al., 1996]. In TATA-Box-haltigen Promotoren ist die Bindung der TATA-Box durch den Faktor TFIID einer der ersten Schritte bei der Bildung des Prä-Initiationskomplexes [Martinez 2002]. Der Faktor TFIID besteht aus dem TATA-Box bindenden Protein (TBP) sowie TAFs (TBP-associated factors), die an der Erkennung des Initiatorelementes (INR) sowie weiterer downstream-Promotorelemente beteiligt sind [Chalkley & Verrijzer 1999]. Nach Bindung von TFIID werden die allgemeinen Faktoren TFIIA und TFIIB gebunden, die den Komplex stabilisieren und die Rekrutierung von TFIIF, Pol II, TFIIE und TFIIH ermöglichen.

Pflanzliche TBP-Proteine binden an die Consensus-DNA-Sequenz TATAWATA [Shahmuradov et al., 2003]. Von einigen TFs, z.B. den TFs GT-1 und HSF, ist bekannt, dass sie mit TBP interagieren können [Le Gourrierec et al., 1999; Reindl & Schöffl 1998]. Der bZIP TF RF2a kann *in vitro* mit dem TBP interagieren [Dai et al., 2003]. Es wird vermutet, dass die Protein-Protein-Interaktion mit anderen TFs die DNA-Bindung des TBP stabilisiert.

Durch die DNA-Bindung des TBP-Proteins und des assoziierten Faktors TFIID wird der Transkriptions-Initiationskomplex an den Promotor rekrutiert und die Position des Transkriptionsstartpunktes (TSS) festgelegt [Zhu et al., 1995].

## 1. Einleitung

### 1.3.2 Der CCAAT-Box bindende Faktor

Ein weiterer allgemeiner TF ist der CCAAT-Box bindende Faktor CBF. Der CBF ist aus drei Untereinheiten aufgebaut (HAP2, 3 und 5), wobei für die jeweiligen Untereinheiten homologe Gene bekannt sind [Edwards et al., 1998]. Die CCAAT-Box tritt nur in einem Teil der Promotoren auf und ist normalerweise im Bereich 60bp-100bp upstream des TSS lokalisiert (siehe Abb. 3) [Mantovani 1999]. Es sind verschiedene gewebespezifisch exprimierte Formen von Untereinheiten des CBF bekannt [Gusmaroli et al., 2001].

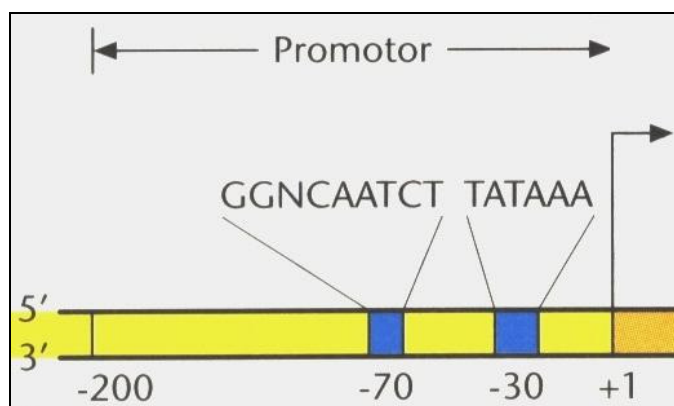


Abb. 3: Position der Bindungsstellen für allgemeine TFs im eukaryontischen Promotor (Abb. verändert nach Straßburger)



#### **1.4 Transkriptionsfaktorbindungsstellen können durch Matrizen repräsentiert werden**

Um die potenziellen Bindungsstellen eines TFs in einer Sequenz detektieren zu können, müssen zunächst die Bindungssequenzen des TFs identifiziert worden sein. Zu diesem Zweck stehen unterschiedliche Systeme zur Verfügung, wie der „Random binding site Selection Assay“ (RBSS) oder das One-Hybrid-System [Ausubel et al., 1988; Ausubel et al., 1998]. Bei dem RBSS wird der entsprechende TF meistens in heterologen Systemen exprimiert, um anschließend mit dem isolierten Protein *in vitro* DNA-Bindungsstudien durchführen zu können. Bei dem One-hybrid-system wird der zu untersuchende Faktor in Hefe exprimiert und die Bindung an eine putative Zielsequenz *in vivo* getestet.

Als Ergebnis solcher Untersuchungen erhält man verschiedene Bindungssequenzen eines Faktors, die häufig einen gemeinsamen Sequenzanteil aufweisen, der relativ unveränderlich und für eine funktionelle Bindung des TFs besonders wichtig ist. Dieser konservierte Teil einer TFBS, der auch häufig in funktionellen TFBS zu finden ist, wird Coresequenz genannt.

Mit den identifizierten Bindungssequenzen wird ein Alignment durchgeführt und die Anzahl der vier verschiedenen auftretenden Nukleotide an jeder Position festgehalten (Abb. 4a). Auf diese Weise erhält man eine Alignmentmatrize, die in die gängigen Matrizen-Suchprogramme wie PATSER [Hertz & Stormo, 1999], MatInspector [Cartharius et al., 2005] oder MATCH <sup>TM</sup> [Kel et al., 2003] eingegeben werden kann.

Für die vorliegende Arbeit wurde das UNIX/Linux-Programm PATSER verwendet; Dieses Programm ist für nicht-kommerzielle Zwecke frei verwendbar, außerdem ist es in der Lage, Sequenzdateien der Größenordnung 30Mb, was der Größe der einzelnen Arabidopsis-Chromosomen entspricht, in ausreichender Geschwindigkeit zu verarbeiten.

Das Suchprogramm wandelt die Alignmentmatrize in eine Positions-Gewichtungsmatrize um, die für das Durchsuchen der Genomsequenz benötigt wird (Abb. 4b).

Anschließend geht das Suchprogramm die Genomsequenz nukleotidweise durch, indem quasi ein Suchfenster der Länge der Matrize über die Sequenz geschoben wird und für jeden neuen Sequenzausschnitt anhand der Matrize ein individueller Score berechnet wird. Dieser Score ist ein Maß für die Übereinstimmung von Matrize

## 1. Einleitung

und dem entsprechenden Bereich der Testsequenz. Er liegt zwischen dem maximalen Score, der sich aus einer möglichst großen Übereinstimmung von Testsequenz und Matrize ergibt, und dem minimalen Score, wenn Testsequenz und Matrize einander möglichst unähnlich sind [Hertz & Stormo, 1999].

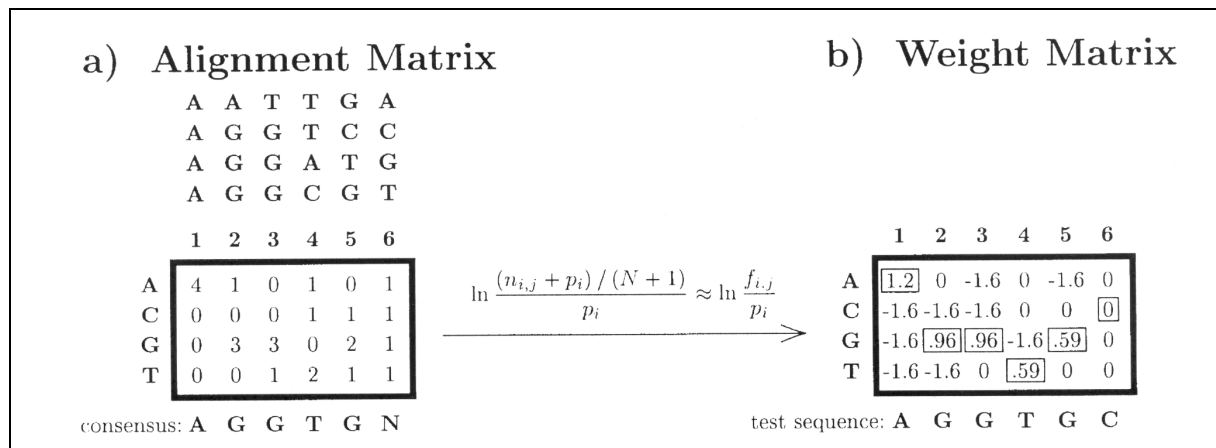


Abb. 4: Vom Alignment der Bindungssequenzen eines TFs zur Matrize [Hertz & Stormo, 1999]

$n_{i,j}$  = Absolute Häufigkeit eines Nukleotides an einer Position

$p_i$  = *a priori* Wahrscheinlichkeit für das Auftreten eines Nukleotides

$N$  = Anzahl der Sequenzen in der Matrize

$f_{i,j}$  = Frequenz eines Nukleotides

Das Programm PATSER legt anhand des Informationsgehaltes der eingegebenen Matrize einen Schwellenwert fest, der als Threshold bezeichnet wird. Ist der Score einer Testsequenz gleich oder größer dem definierten Schwellenwert, so wird die aktuelle Sequenz als potenzielle Bindungsstelle ausgegeben. Die ausgegebenen Parameter für jede potenzielle Bindungsstelle sind: Name der Sequenz, Position, Orientierung und individueller Score.

### 1.5 Die Bindungsstellen interagierender TFs können kombinatorische Elemente bilden

Einige TFs benötigen für eine funktionelle DNA-Bindung die Interaktion mit einem weiteren TF [Singh 1998]. In diesen Fällen bildet sich ein Dimer aus den beiden Proteinen, wobei es sich je nach TF um ein Homodimer handeln kann (z.B. bei den bZIP-TFs) oder ein Heterodimer, das aus verschiedenen TFs aufgebaut ist (z.B. MYC und MYB). Die Interaktionen, an denen ein TF teilnimmt, sind durch seine Interaktionsdomäne festgelegt und sind spezifisch für jeden TF. Beispiele für

kombinatorische Interaktion sind bei der Auxin-vermittelten Induktion von Genen durch die ARF-TFs (Auxin response factors) bekannt [Ulmasov et al., 1999] oder bei der durch Dehydratation induzierten Genexpression [Abe et al., 1997], die teilweise durch interagierende MYC und MYB-Faktoren vermittelt wird. Ebenfalls Zielgene von interagierenden MYC / MYB-TFs sind einige Gene der Anthocyan-Biosynthese, wie z.B. das BANYULS-Gen, für das durch eine Promotor-Deletionsanalyse gezeigt wurde, dass es durch die TFs TT8 (MYB) und TT2 (MYC) synergistisch reguliert wird [Baudry et al., 2004].

Für zwei DNA-gebundene Proteine, die miteinander in Protein-Protein-Wechselwirkung treten, müssen sich die Bindungsstellen in der Regel in räumlicher Nähe auf der DNA befinden bzw. durch Reorganisation des Chromatins in räumliche Nähe gebracht werden. In einigen Fällen müssen die TFBS sehr eng benachbart sein, wie im Falle des kombinatorischen as-1-Elementes, das aus zwei bZIP-Bindungsstellen aufgebaut ist, deren erste Nukleotide nicht mehr als 15bp voneinander entfernt sind [Krawczyk et al., 2002].

Kombinatorische Elemente können aber auch aus mehr als nur zwei TFBS bestehen und erheblich länger sein als das as-1-Element, wie z.B. im Falle der kombinatorischen MADS-Elemente. Die TFs der MADS-Klasse sind zu vielfältigen Interaktionen mit anderen MADS-Faktoren fähig, und die resultierenden kombinatorischen Elemente sind möglicherweise aus drei MADS-Bindungsstellen aufgebaut [Egea-Cortines et al., 1999]. Die genaue Art der Interaktion ist bis heute noch nicht bekannt, es wird jedoch die Existenz eines ternären Komplexes aus drei unterschiedlichen MADS-Proteinen diskutiert, der in der Blütenentwicklung von *Arabidopsis thaliana* involviert ist .

### **1.6 Colokalisationsanalyse von TFBS kann Hinweise auf potenzielle kombinatorische Elemente geben**

Wie unter Abschnitt 1.5 dargelegt, sind Protein-Protein-Interaktionen zwischen TFs ein wichtiger Parameter bei der Spezifizierung von Protein-DNA-Interaktionen. Es erscheint nahe liegend, dass für die Unterscheidung zwischen funktionellen TFBS und nicht-funktionellen TFBS nicht nur die Sequenz der Bindungsstelle bedeutsam sein kann. Dies lässt sich besonders gut am Beispiel der bZIP-TFs illustrieren.

Die Erkennungssequenz der meisten in Pflanzen bekannten bZIP-TFs ist ca. 10 Nukleotide lang und besitzt die stark konservierte Kernsequenz ACGT mit jeweils drei flankierenden Nukleotiden, die eine deutlich schwächere Konservierung zeigen [Izawa et al., 1993]. Allein die Kernsequenz tritt bei einer angenommenen Gleichverteilung der vier Nukleotide statistisch alle 256bp auf, und die G-Box, die eine potenzielle Bindungsstelle für mehrere bZIP-Faktoren darstellt, tritt etwa alle 4kb auf. Wie unterscheiden verschiedene TFs zwischen tatsächlichen Zielsequenzen im regulatorischen Bereich von Genen und homologen Sequenzen im codierenden Bereich? Wie wird festgelegt, dass eine bZIP-Site in einem Gewebe funktionell ist und in einem anderen Gewebe keine regulatorische Rolle erfüllt? Zum einen spielen sicherlich epigenetische Faktoren wie DNA-Methylierung eine wichtige Rolle bei der Erkennung funktioneller TFBS in Promotorbereichen [Finnegan et al., 2000]. Zum anderen sind jedoch auch Interaktionen zwischen zwei TFs ein wichtiger Parameter bei der Spezifizierung der TF-DNA-Interaktion. Die räumliche Distanz zwischen zwei TFBS spielt dabei eine wichtige Rolle. Die Colokalisation zweier TFBS kann ein Hinweis auf Protein-Protein-Interaktion der korrespondierenden TFs sein. Aus den Untersuchungen von Krawczyk et al. [2002] ist bekannt, dass die Distanz zwischen zwei potenziell synergistischen TFBS für das Zustandekommen einer funktionellen TF-DNA-Bindung am as-1-Element essentiell ist [Krawczyk et al., 2002]. Neuere Untersuchungen belegen, dass auch kombinatorisch aufgebaute ABREs aus zwei G-Box homologen Sequenzen nur bis zu einer bestimmten Distanz (ca. 150bp) zueinander funktionell sind [Zhang et al., 2005].

Es wurde nachgewiesen, dass die TFBS synergistischer TFs im Säugetiergenom häufiger colokalisieren als die TFBS von TFs, die keine gemeinsamen Zielgene aufweisen [Hannenhalli & Levy, 2002]. Im Rahmen dieser Arbeit sollte der methodische Ansatz der Colokalisationsanalyse von TFBS im genomischen Kontext auch für die Analyse des *Arabidopsis*-Genoms umgesetzt und genutzt werden (siehe Abschnitt 1.9).

### **1.7 Motivation zur Erstellung von AthaMap, einer genomweiten Datenbank *cis*-regulatorischer Elemente in *Arabidopsis thaliana***

Ziel dieser Arbeit war es, eine genomweite Datenbank für *cis*-regulatorische Elemente im *Arabidopsis thaliana* Genom zu schaffen, die möglichst umfassend ist und genug Flexibilität für zukünftige Erweiterungen bietet. Da funktionelle *cis*-regulatorische Elemente nicht nur in upstream-Bereichen von Genen auftreten, sondern auch z.B. in Introns und anderen nichtcodierenden Bereichen, muss die gesamte *Arabidopsis*-Genomsequenz miteinbezogen werden. Die Bindungsspezifität eines TFs kann sich bei kombinatorischer Interaktion mit einem anderen TF ändern, aus diesem Grunde sollten keine Consensussequenzen zur Suche nach potenziellen TFBS verwendet werden, sondern publizierte Alignmentmatrizen von pflanzlichen TFs bzw. selbst erstellte Matrizen aus mehreren einzelnen publizierten Bindungsstellen pflanzlicher TFs. Um spezifische Analysen durchführen zu können, sollten Funktionen implementiert werden, die eine Reduktion der in AthaMap angezeigten TFBS auf höher konservierte TFBS ermöglichen.

Ein weiteres Ziel bei der Erstellung von AthaMap war die möglichst genaue Annotation putativer Bindungsstellen des TATA-Box bindenden Proteins. Wie bereits unter Abschnitt 1.3 erwähnt, nimmt der Faktor TBP durch seine Aufgaben bei der Rekrutierung des Transkriptions-Initiationskomplexes zum Corepromotor eine Sonderstellung unter den TFs ein. Ferner bietet die Annotation putativ funktioneller TATA-Boxen die Möglichkeit, durch Colokalisationsanalysen potenzielle Interaktionspartner des TBP zu identifizieren.

Da die Bedeutung von kombinatorischen Elementen für die Spezifität der Genregulation immer stärker deutlich wird, sollten potenzielle kombinatorische Elemente basierend auf bekannte TF-Interaktion genauso wie die einzelnen TFBS in die AthaMap-Datenbank aufgenommen werden.

Basierend auf der Beobachtung, dass TF-Interaktion häufig zwischen räumlich eng benachbarten TFs abläuft, sollte eine Colokalisationsfunktion in die Datenbankoberfläche integriert werden. Diese Funktion soll die Analyse colokalisierender TFBS sowie die Identifikation von neuen kombinatorischen Elementen anhand benutzerdefinierter Parameter ermöglichen. Der Benutzer soll die potenziell beteiligten TFs wählen können sowie den maximalen und minimalen Abstand zwischen beiden TFBS und den Konservierungsgrad (Score) der in die Analyse mit einbezogenen TFBS. Auf diese Weise wäre es möglich, experimentell

gewonnene Daten über TF-Interaktionen *in silico* weiter zu analysieren und auf diese Weise potenzielle Zielgene eines kombinatorischen Elementes zu identifizieren. Die oben beschriebenen Funktionen sollten in die AthaMap-Datenbank implementiert werden und am Beispiel von Abscisinsäure-responsiven Genen getestet werden.

### **1.8 Weitere Datenbanken für die Analyse der pflanzlichen Genregulation und *cis*-regulatorischer Elementen**

Um pflanzliche Genregulation untersuchen zu können, stehen zwar mehrere Datenbanken und Internet-Ressourcen verfügbar, die aber nicht für genomweite Analysen eingesetzt werden können und individuelle Vor- und Nachteile haben. Eine der bedeutsamsten Informationsquellen für TF-Daten ist die TRANSFAC-Datenbank [Matys et al., 2003] (<http://www.gene-regulation.com/>). Die TRANSFAC-Datenbank enthält gesammelt publizierte Daten über tierische, pflanzliche und bakterielle TFs sowie TFs, die in Pilzen vorkommen. In den Fällen, in denen Bindungssequenzen für die TFs bekannt sind, wurden diese in der TRANSFAC-Datenbank annotiert und, wenn möglich, Alignmentmatrizen daraus erstellt. Es ist ferner möglich, eigene Sequenzen einzugeben und diese anhand von Matrizen (MATCH) oder Consensussequenzen (PATCH) nach potenziellen TFBS durchsuchen zu lassen. Im Falle der Matrizensuche ist es möglich, eine vorher vom Benutzer definierte Auswahl von Matrizen für die Suche zu verwenden. TRANSFAC war zu Beginn dieser Arbeit die einzige Ressource, bei der TFBS durch Matrizen repräsentiert werden, allerdings bietet TRANSFAC nicht die Möglichkeit eines genomischen Matrizenscreenings an.

Eine weitere Ressource für regulatorische Sequenzen in Pflanzen ist die PlantCARE-Datenbank (<http://intra.psb.ugent.be:8080/PlantCARE/>) [Lescot et al., 2002]. Hier kann mithilfe einer Suchfunktion nach bereits bekannten *cis*-regulatorischen Elementen in eigenen Sequenzen gesucht werden. Es ist allerdings nicht möglich, ganze Chromosomen nach TFBS zu durchsuchen. Die PLANTCare-Datenbank enthält nur relativ wenige (ca. 400) bekannte TFBS aus verschiedenen Spezies, die durch Consensussequenzen repräsentiert werden.

Eine ähnliche Ressource stellt PLACE (plant *cis*-acting elements) dar (<http://www.dna.affrc.go.jp/PLACE/signalscan.html>) [Higo et al., 1999]. In der PLACE Datenbank sind pflanzliche *cis*-regulatorische Elemente gespeichert, nach denen der

## 1. Einleitung

Benutzer seine eigene Sequenz durchsuchen lassen kann. Die Suche wird mit Consensussequenzen durchgeführt, und die zu durchsuchende Sequenz darf nicht länger als 4.356 Nukleotide sein, da sonst das Ergebnis leer ist. Außerdem kann auf der PLACE-Seite eine Homologie-Suche mit eigenen Sequenzen durchgeführt werden.

Die Datenbanken ATCISDB und ATTFDB sind miteinander verbunden und komplementieren sich gegenseitig [Davuluri et al., 2003]. In der Datenbank ATCISDB (<http://arabidopsis.med.ohio-state.edu/AtcisDB/>) sind potenzielle TFBS in putativen Genpromotoren von *Arabidopsis thaliana* gespeichert. Als putativer Promotor wird hier der upstream-Bereich eines Gens bis zur nächsten codierenden Region bzw. bis zu einer Länge von max. 3kb betrachtet, nichtcodierende downstream-Sequenzen (z.B. Introns) werden generell nicht miteinbezogen. Verbunden mit der ATCISDB ist die Datenbank ATTFDB, die Daten über pflanzliche Transkriptionsfaktoren enthält (<http://arabidopsis.med.ohio-state.edu/AtTFDB/>).

Für jedes annotierte Gen in ATCISDB existieren Links zu externen Datenbanken. Der Gesamtumfang der Daten in ATCISDB ist nicht sichtbar. Ein weiterer Nachteil ist, dass potenzielle TFBS zum Teil durch die Suche nach Consensussequenzen identifiziert wurden und daher kaum Variabilität in den Bindungsstellen zugelassen wird. Außerdem wurden in ATCISDB nur potenzielle TFBS in upstream-Regionen annotiert, obgleich bekannt ist, dass auch Introns regulatorische Wirkung besitzen können. Da immerhin die upstream-Regionen aller *Arabidopsis*-Gene erfasst wurden, bilden ATTFDB und ATCISDB eine nützliche Ressource, die durch Integration des gesamten *Arabidopsis*-Genoms noch weiter verbessert werden könnte.

Eine sehr vielseitige Ressource stellt die TAIR-Datenbank dar ([www.arabidopsis.org](http://www.arabidopsis.org)) [Rhee et al., 2003]. Sie beschäftigt sich zwar nicht primär mit Genregulation oder regulatorischen Elementen, dennoch werden die für diese Arbeit relevanten Funktionen und Inhalte hier kurz zusammengefasst. Zunächst ist es mithilfe der TAIR-Suchfunktion möglich, allgemeine Informationen zu jedem annotierten *Arabidopsis*-Gen sowie den eindeutigen AGI-Bezeichner (*Arabidopsis* genome initiative) anzeigen zu lassen.

In dem „Sequence viewer“ der TAIR-Datenbank wird die gesamte *Arabidopsis*-Genomsequenz mitsamt der Genstruktur dargestellt. Auch alternative Transkripte eines Gens werden hier dargestellt und farblich unterschieden. Der anzuzeigende

Genombereich kann vom Benutzer entweder manuell durch Eingabe einer Position oder durch Eingabe eines Gennamens gewählt werden.

Eine essentielle Ressource, die sich mit Microarrayexperimenten befasst, ist der Genevestigator ([www.genevestigator.ethz.ch](http://www.genevestigator.ethz.ch)) [Zimmermann et al., 2004]. Mithilfe der „Meta-Analyzer“-Funktion der Genevestigator-Datenbank kann der Transkriptionsstatus von 50 Genen gleichzeitig in verschiedenen Experimenten abgefragt werden. Das Spektrum der gespeicherten Ergebnisse umfasst z.B. biotischen und abiotischen Stress, Hormonwirkung und verschiedene Lichtsituationen. Gleichmaßen ist es mit der Meta-Analyzer-Funktion möglich, Informationen zur gewebe- oder entwicklungsspezifischen Gentranskription eines Gens zu erhalten.

Genevestigator ist zurzeit die vollständigste und schnellste Webressource für Microarray-Experimente bei Pflanzen.

Weitere Ressourcen im Zusammenhang mit pflanzlicher Genregulation sind z.B. die „Eukaryotic promotor database“ (<http://www.epd.isb-sib.ch/>) [Schmid et al., 2004] sowie die „PlantProm“-Datenbank (<http://mendel.cs.rhul.ac.uk/mendel.php>) [Shahmuradov et al., 2003].

Beim Vergleich der hier aufgezählten Ressourcen wird deutlich, dass bei allen oben aufgezählten Datenbanken lediglich Consensussequenzen zur Beschreibung der regulatorischen Elemente verwendet werden (Ausnahme: TRANSFAC) und kein vollständiges Pflanzengenom analysiert werden kann. Außerdem existierte derzeit noch keine Datenbank, in der potenzielle regulatorische Elemente genomweit annotiert sind. Da *cis*-regulatorische Elemente im genomischen Kontext interagieren können, sollte eine Ressource geschaffen werden, die diesem Umstand Rechnung trägt und eine genomweite Analyse in *Arabidopsis thaliana* ermöglicht.

### **1.9 Aufbau Abscisinsäure-responsiver Elemente (ABREs)**

Von ABREs wird angenommen, dass sie generell kombinatorischen Charakter besitzen [Choi et al., 2000]. Das Phytohormon ABA ist in verschiedenen Entwicklungsprozessen der Pflanze, wie z.B. der Samenreifeung oder der Keimung, involviert [Leung & Giraudat 1998]. Darüber hinaus ist ABA als stressinduziertes Phytohormon in verschiedene Adaptionsprozesse involviert, unter anderem bei



Anpassungen an Dehydratation, Salzstress oder Kälte [Shinozaki & Yamaguchi-Shinozaki, 1996].

Bei der Regulation der Abscisinsäure-induzierten Genexpression sind verschiedene Abscisinsäure-responsive Elemente (ABREs) bekannt, die unterschiedlich aufgebaut sind. Viele dieser Elemente beinhalten die Sequenz ACGT, unter ihnen das ABRE des Em-Gens aus Weizen, das als Kernsequenz eine perfekte G-Box besitzt (CACGTG) [Choi et al., 2000]. Mittlerweile sind eine Reihe von ABREs bekannt, die auch die ACGT-Coresequenz besitzen, aber keine perfekten G-Boxen darstellen. Ein anderer Typ von ABRE enthält das so genannte „Coupling element“, das eine abgewandelte Sequenz (CGCGTG) enthält, aber in *Arabidopsis*-ABREs selten vorkommt [Zhang et al., 2005]. Generell wird davon ausgegangen, dass ABREs nicht aus nur einer TF-Bindungsstelle bestehen, sondern kombinatorischen Charakter besitzen [Choi et al., 2000]. Neuere Untersuchungen belegen, dass eine Kombination von G-Boxen überrepräsentiert häufig in ABA-induzierbaren Genen auftritt [Zhang et al., 2005]. Diese ABRE-Module treten größtenteils im Bereich bis - 100bp relativ zum nächsten TSS auf, können aber in Einzelfällen auch in der 5'UTR lokalisiert sein. Die einzelnen Elemente des Moduls, G-Box und CE, treten bevorzugt in einem Abstand bis zu 100bp voneinander entfernt auf, in selteneren Fällen auch bis zu 150bp voneinander entfernt. Auf Basis dieser Daten wurden von Zhang et al. potenziell ABA-responsive Zielgene identifiziert und mithilfe von RT-PCR gezeigt, dass diese in 67,5% der überprüften Fälle tatsächlich ABA-induzierbar sind [Zhang et al., 2005].

Verschiedene Transkriptionsfaktoren, die eine entwicklungsbedingte ABA-Aktivität vermitteln können, wurden mittlerweile untersucht [Kagaya et al., 2005] [Bensmihen et al., 2005]. Über die TFs, die spezifisch in die Regulation der ABA-induzierten Stressadaptation involviert sind, ist bisher weniger bekannt. Eine Gruppe von ABA-streßinduzierten TFs sind die ABFs [Choi et al., 2000], die zur Familie der bZIP-TFs gehören. In AthaMap wurden die potenziellen Bindungsstellen des TFs ABF1 annotiert, der in der Lage ist, an ABREs verschiedener Typen zu binden, darunter ABREs mit und ohne ACGT-Coresequenz.

Da von mehreren Mitgliedern der bZIP-Familie bekannt ist, dass sie mit dem TBP interagieren können [Dai et al., 2003], wurden Colokalisationsanalysen von ABF1-Bindungsstellen und der TATA-Box durchgeführt. Es wurde untersucht, ob kombinatorische Elemente aus ABF1-Bindungsstelle und TATA-Boxen bei der

## 1. Einleitung

Induktion von ABA-responsiven Genen eine Rolle spielen. Zum Vergleich mit experimentellen Daten wurden die ABA-induzierten Genexpressionsprofile aus der Genevestigator-Datenbank [Zimmermann et al., 2004] verwendet.

## 2. Material und Methoden

### 2.1 Genom-Daten

Für die aktuelle Version der AthaMap Datenbank [Steffens et al., 2005] wurde die genomische *Arabidopsis* Sequenz benutzt, die vom „Institute for genome research“ (TIGR; <http://www.tigr.org/> ; [Haas et al., 2005]) zur Verfügung gestellt wird (Version vom 21.01.2004). Bei der Genomversion von TIGR liegen die Sequenz jedes Chromosoms und die zugehörige Annotation gemeinsam in einer XML-Datei vor. Mithilfe eines PERL-Skriptes (Claudia Galuschka, persönliche Mitteilung) wurde die reine Genomsequenz von den Annotationsdaten getrennt und in Form von Textdateien gespeichert. Diese Textdateien, die jeweils die Sequenz eines Chromosoms repräsentieren, wurden zur Identifizierung potenzieller TFBS verwendet.

### 2.2 Material für Matrizenscreenings

Für das genomweite Screening benötigt das verwendete Suchprogramm PATSER Alignmentmatrizen von TFBS als Template. Im Folgenden sind sämtliche TFs aufgeführt, deren Matrizen verwendet wurden. Die einzelnen TF-Bindungssequenzen für die Generierung der Matrizen bzw. die Matrizen selbst wurden aus den Referenzen extrahiert und sind im Anhang unter Abschnitt 7.1 aufgeführt. Tabelle 1 zeigt eine Auflistung sämtlicher TFs, deren TFBS mithilfe von Alignmentmatrizen detektiert und in der AthaMap-Datenbank gespeichert wurden.

## 2. Material und Methoden

Tabelle 1: Eine Auflistung der TFs, deren potenzielle Bindungsstellen in AthaMap gespeichert vorliegen.

TF	Familie	Spezies	Referenz für Alignmentmatrize
ABF1	bZIP	<i>Arabidopsis thaliana</i>	[Choi et al., 2000]
bZIP910[1]	bZIP	<i>Antirrhinum majus</i>	[Martinez-Garcia et al., 1998]
bZIP910[2]	bZIP	<i>Antirrhinum majus</i>	[Martinez-Garcia et al., 1998]
bZIP911[1]	bZIP	<i>Antirrhinum majus</i>	[Martinez-Garcia et al., 1998]
bZIP911[2]	bZIP	<i>Antirrhinum majus</i>	[Martinez-Garcia et al., 1998]
TGA1	bZIP	<i>Arabidopsis thaliana</i>	[Schindler et al., 1992]
TGA1a	bZIP	<i>Nicotiana tabacum</i>	[Izawa et al., 1993]
O2	bZIP	<i>Zea mays</i>	[Izawa et al., 1993]
PIF3[1]	bHLH	<i>Arabidopsis thaliana</i>	[Martinez-Garcia et al., 2000]
PIF3[2]	bHLH	<i>Arabidopsis thaliana</i>	[Martinez-Garcia et al., 2000]
DOF2	DOF	<i>Zea mays</i>	[Yanagisawa & Schmidt, 1999]
AG	MADS	<i>Arabidopsis thaliana</i>	[Huang et al., 1993]
AGL3	MADS	<i>Arabidopsis thaliana</i>	[Huang et al., 1995]
AGL15	MADS	<i>Arabidopsis thaliana</i>	[Tang & Perry, 2003]
ABI4[1]	AP2/EREBP	<i>Zea mays</i>	[Niu et al., 2002]
ABI4[2]	AP2/EREBP	<i>Zea mays</i>	[Niu et al., 2002]
ANT	AP2/EREBP	<i>Arabidopsis thaliana</i>	[Nole-Wilson & Krizek, 2000]
RAV1[1]	AP2/EREBP	<i>Arabidopsis thaliana</i>	[Kagaya et al., 1999]
RAV1[2]	AP2/EREBP	<i>Arabidopsis thaliana</i>	[Kagaya et al., 1999]
TEIL	AP2/EREBP	<i>Nicotiana tabacum</i>	[Kosugi & Ohashi, 2000]
AtMYB15	MYB	<i>Arabidopsis thaliana</i>	[Romero et al., 1998]
AtMYB77	MYB	<i>Arabidopsis thaliana</i>	[Romero et al., 1998]
AtMYB84	MYB	<i>Arabidopsis thaliana</i>	[Romero et al., 1998]
CDC5	MYB	<i>Arabidopsis thaliana</i>	[Hirayama & Shinozaki, 1996]
GAMYB	MYB	<i>Hordeum vulgare</i>	[Gubler et al., 1999]
MYB.PH3[1]	MYB	<i>Petunia hybrida</i>	[Solano et al., 1995]
MYB.PH3[1]	MYB	<i>Petunia hybrida</i>	[Solano et al., 1995]
P	MYB	<i>Zea mays</i>	[Grotewold et al., 1994]
GT1	Trihelix	diverse Spezies	[Ouwerkerk et al., 1999]
PCF2	TCP	<i>Oryza sativa</i>	[Kosugi & Ohashi 2002]
PCF5	TCP	<i>Oryza sativa</i>	[Kosugi & Ohashi 2002]
HVH21	HD-Kn	<i>Hordeum vulgare</i>	[Krusell et al., 1997]
ALFIN1	HD-PHD	<i>Medicago sativa</i>	[Bastola et al., 1998]
ATHB1	HD-ZIP	<i>Arabidopsis thaliana</i>	[Sessa et al. 1993]
ATHB5	HD-ZIP	<i>Arabidopsis thaliana</i>	[Johannesson et al., 2001]
ATHB9	HD-ZIP	<i>Arabidopsis thaliana</i>	[Johannesson et al., 2001]
HAHB4	HD-ZIP	<i>Helianthus annuus</i>	[Palena et al., 1999]
AGP1	GATA	<i>Nicotiana tabacum</i>	[Sugimoto et al., 2003]
ZAP1	WRKY	<i>Arabidopsis thaliana</i>	[de Pater et al., 1996]
ID1	C2H2 (Zn)	<i>Zea mays</i>	[Kozaki et al., 2004]
TBP		diverse Spezies	[Shahmuradov et al., 2003]
CBF		diverse Spezies	[Shahmuradov et al., 2003]

### 2.3 Identifizierung potenzieller TFBS

Zur Identifizierung potenzieller TFBS wurde das lokal installierte Programm PATSER mit publizierten Alignmentmatrizen verwendet [Hertz & Stormo 1999; Steffens et al., 2004; Steffens et al., 2005]. PATSER wurde mitsamt der Dokumentation von der FTP-Site des Autors (<ftp://beagle.colorado.edu/pub/consensus>) heruntergeladen. Als Basissystem zum Betrieb von PATSER dient ein S.U.S.E. LINUX-Betriebssystem der Version 8.1 auf einem herkömmlichen PC. Die Bedienung von PATSER erfolgt über

## 2. Material und Methoden

eine Kommandozeile in einer LINUX Shell, wie in der Dokumentation zu PATSER beschrieben. Als alternative Bedienoberfläche für PATSER wurde ein Webinterface von Martin Schindler entwickelt, das die Bedienung wesentlich vereinfacht (zugänglich unter: <http://134.169.87.92/patser.php> bzw. <http://134.169.87.92/patser2zip.php>). Nachfolgend wird die Bedienung kurz erläutert, soweit die Bedienoberfläche nicht selbsterklärend ist.

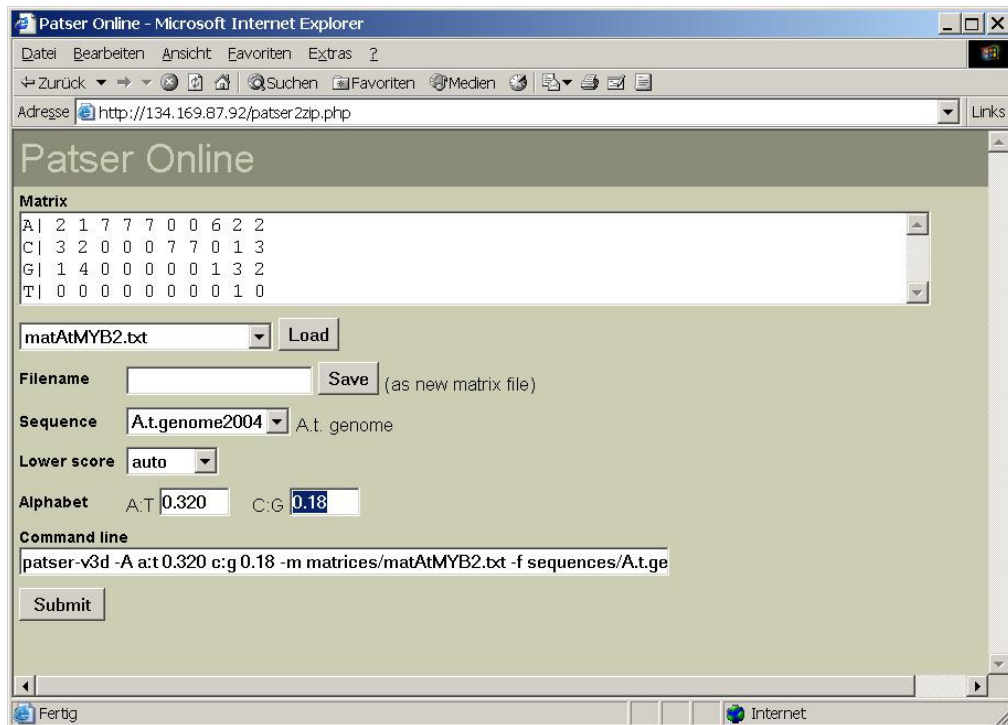


Abb. 5: PATSER Bedienoberfläche

Um mit einer neuen Matrize ein Screening durchzuführen, muss diese zunächst im PATSER Verzeichnis gespeichert werden. Zu diesem Zweck fügt man die Alignmentmatrize in das Textfenster ganz oben auf der Seite ein (Abb. 5), gibt in dem Textfeld, das mit „filename“ bezeichnet ist, einen Namen für die Matrize an (gewöhnlich „Name.txt“) und speichert diese durch Druck auf den „Save“ Knopf. Die Matrize (matAtMYB2.txt, Abb. 5) liegt nun gespeichert vor und kann im Auswahlfenster unterhalb des Textfensters für ein Screening ausgewählt werden. In dem genannten Auswahlfenster sind die Namen aller bereits vorhandenen Matrizen aufgelistet. Für eine Matrizensuche mit einer schon gespeicherten Matrize muss lediglich eine Matrize aus der Liste ausgewählt werden. Bei Druck auf den „Load“-Knopf wird die Matrize dargestellt (Abb. 5 AtMYB2). Dies ist jedoch zum Screenen mit der Matrize nicht unbedingt nötig.

Als nächstes muss die zu durchsuchende Sequenz im „sequence“ Auswahlfenster ausgewählt werden. Derzeit stehen zwei Genomversionen von *Arabidopsis thaliana* sowie einige Testsequenzen zur Verfügung. Hierbei handelt es sich um die aktuelle Genomversion von TIGR (Abb. 5, A.t.genome2004) und eine ältere Genomversion von TAIR (A.t.genome2003). Neue Testsequenzen können eingespielt werden, indem sie einfach in das „Sequence“-Unterverzeichnis von PATSER kopiert werden (nicht gezeigt). Diese Sequenzen erscheinen beim nächsten Öffnen der Bedienoberfläche automatisch im entsprechenden Auswahlfenster.

Nach Auswahl der Sequenz kann bestimmt werden, ob der Threshold-Score für das Screening automatisch von PATSER festgelegt werden soll (Lower score: „Auto“, Abb. 5) oder ob ein Score manuell vorgegeben werden soll. Wählt man in dem Auswahlfenster „Lower score“ die Option „manuell“, so wird ein Textfenster eingeblendet, in das ein Threshold-Score für das Screening eingegeben werden kann. Bei manueller Eingabe eines Thresholds müssen Dezimalstellen durch einen Punkt, kein Komma, abgetrennt werden.

Zur korrekten Berechnung eines automatischen Threshold-Scores wird die Nukleotid-Komposition des Genoms herangezogen. Es ist daher darauf zu achten, dass der angegebene GC-Gehalt des Genoms in den Textfenstern „A:T“ und „C:G“ mit dem GC-Gehalt der ausgewählten zu durchsuchenden Sequenz übereinstimmt (siehe Abb. 5, „Command line“). Für die aktuelle Genomversion von 2004 liegt der A:T-Gehalt bei 0.32 und der GC-Gehalt entsprechend bei 0.18.

Hat man alle oben erwähnten Parameter eingegeben bzw. ihre Richtigkeit überprüft, so kann die Analyse mit einem Druck auf den „Submit“-Knopf gestartet werden.

Für genomweite Analysen sollte die PATSER2ZIP-Bedienoberfläche (<http://134.169.87.92/patser2zip.php>) verwendet werden, die das Ergebnis als Datei direkt auf eine zuvor eingelegte ZIP-Diskette des PCs schreibt. Der Name der Ergebnisdatei wird automatisch vom PATSER-Webinterface generiert, er setzt sich aus dem Namen der gewählten Matrize und dem Systemdatum zusammen, zu dem die Analyse gestartet wurde. Sobald das Screening beendet ist, erscheint eine Meldung auf dem Bildschirm des Rechners.

Die Ergebnisse der einfachen PATSER-Bedienoberfläche (<http://134.169.87.92/patser.php>) werden direkt auf dem Bildschirm in Echtzeit ausgegeben. Mithilfe dieser Funktion ist es möglich, Testscreenings mit neuen Matrizen durchzuführen, um Informationen über die Matrizenparameter wie

Threshold-Score, max. Score etc. zu erhalten. Genomweite Analysen sind mit dieser PATSER-Version nicht ratsam, da es aufgrund langer Rechenzeiten bei diesen Analysen zu einem Timeout kommen kann.

### 2.4 Import von TFBS in die AthaMap Datenbank

Als Tool für den Datenimport wurde das Programm „TFBSitesTools“ benutzt, das von Lorenz Bülow zur Verfügung gestellt wurde (persönliche Mitteilung). Dieses Programm wurde in Visual Basic 6.0 entwickelt und läuft auf jedem herkömmlichen Windows-PC ab Win95. Die TFBSitesTools ermöglichen sowohl den Import als auch die Analyse von potenziellen TFBS in der AthaMap Datenbank. Nachfolgend wird die Bedienung des Programms kurz erläutert, soweit sie nicht selbsterklärend ist. Nach dem Starten des Programmes erscheint zunächst ein Auswahlbildschirm (Abb. 6), auf dem man zwischen verschiedenen Programmfunktionen wählen kann.

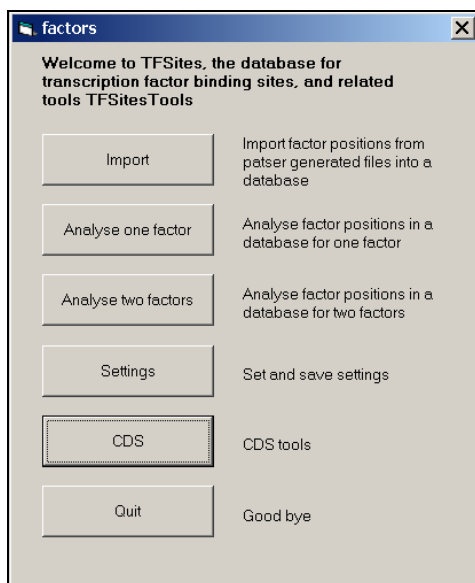


Abb. 6: TFBSitesTools Auswahlfenster

Wichtig ist zunächst, durch Druck auf den „Settings“-Knopf die Einstellungen anzeigen zu lassen und zu überprüfen (Abb. 7, nächste Seite). Hier sind der Name des SQL-Servers (HEHL5), der Name der zu bearbeitenden Datenbank (AthaMap2004work) sowie der Benutzername (factor) einzutragen. Der Server sowie der für den Import verwendete PC müssen in einem lokalen Netzwerk verbunden sein.

## 2. Material und Methoden

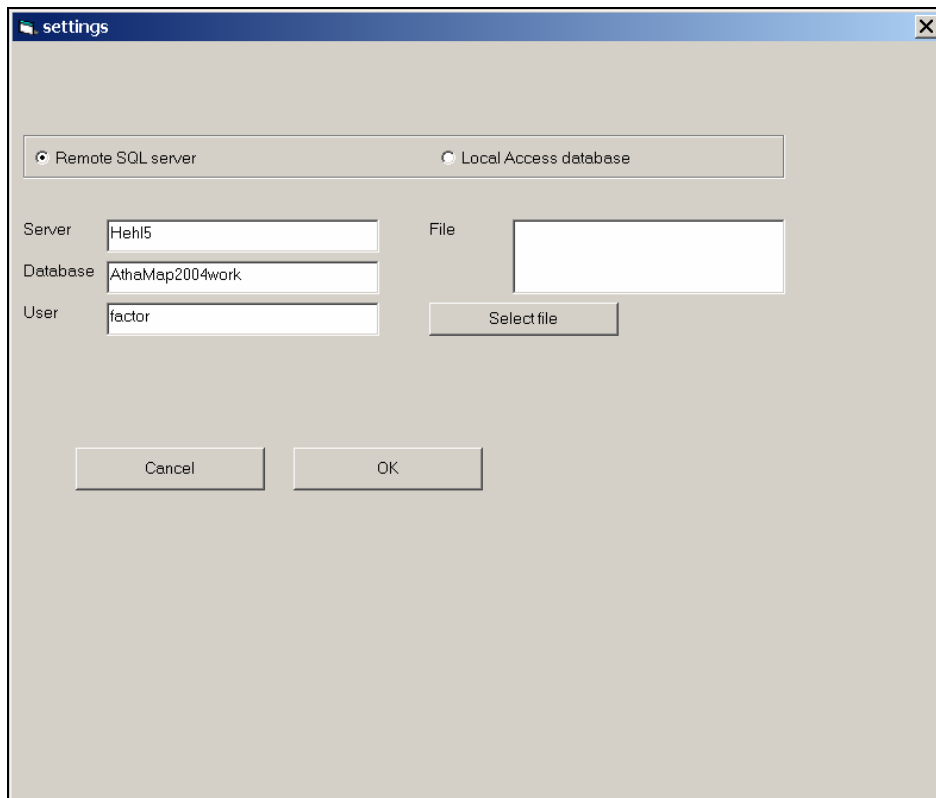


Abb. 7: TFBSitesTools Einstellungsfenster

Es ist darauf zu achten, dass der gewählte Benutzer die nötigen Rechte (Schreibrecht, Erstellung von Tabellen etc.) für die gewählte Datenbank besitzt. Hat er diese nicht, so müssen die entsprechenden Rechte mithilfe eines Administrationstools (z.B. MS SQL Server Enterprise Manager) gesetzt werden. Durch Druck auf den „Import“-Knopf (Abb. 6) wird das Importfenster des Programms geöffnet (Abb. 8).

Zu allererst wird das Format der PATSER-Ergebnisdatei durch den Radio-Button oben rechts ausgewählt, in unserem Fall „old format“ für die Ergebnisse unserer unter Abschnitt 2.3 beschriebenen PATSER-Bedienoberfläche (Abb. 8).

Die Option „new format“ ist auf das Ausgabeformat des bei den RSA-Tools zugänglichen PATSER-Programms abgestimmt [van Helden, 2003].

Nun wird die ZIP-Diskette, auf der das PATSER-Screening gespeichert wurde, eingelegt. Anschließend wird durch Druck auf den „Select file“-Knopf ein Auswahlfenster für die Ergebnisdatei ausgewählt, der Dateiname wird im Textfenster „File“ neben dem Knopf angezeigt, der Anfang der Datei wird zur Kontrolle im Textfenster darunter („Content“) dargestellt. Zusätzlich wird in der „Comment“-Zeile angezeigt, ob die gewählte Datei das richtige Format hat oder nicht.



## 2. Material und Methoden

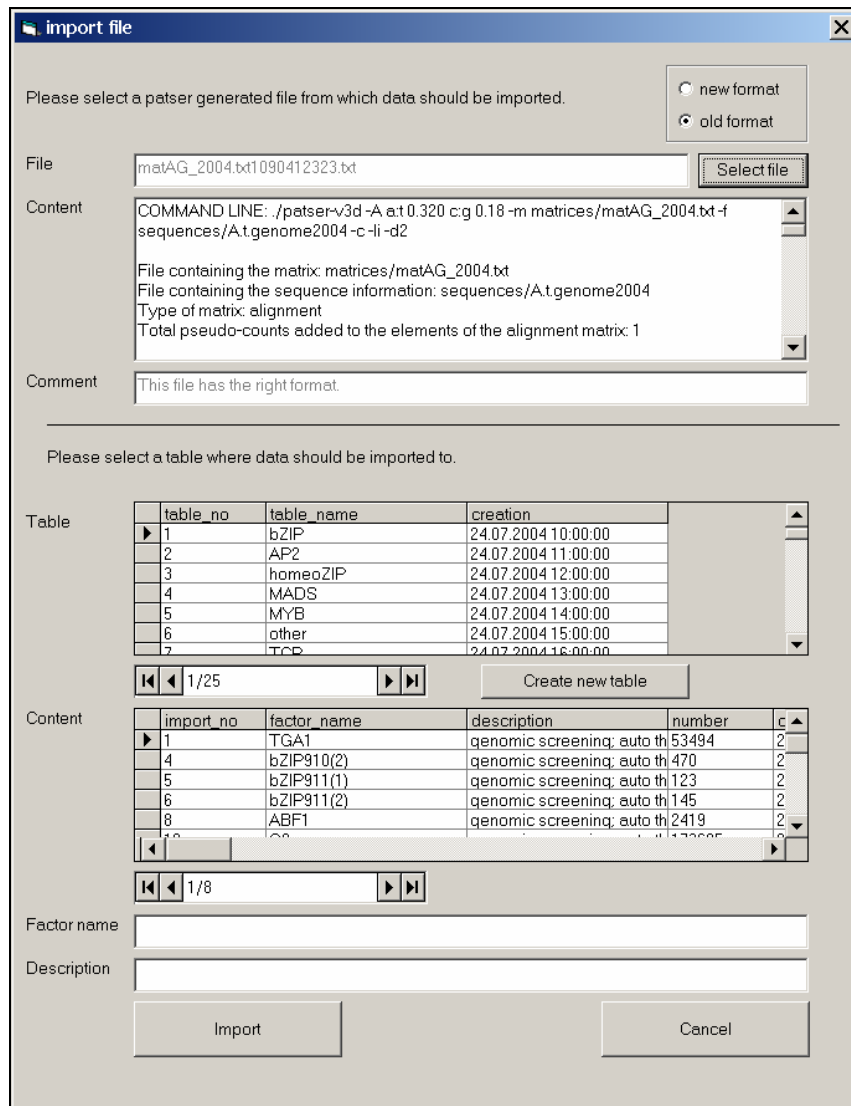


Abb. 8: TFBSitesTools Importfenster

Vor dem Importieren der potenziellen TF-Bindungsstellen muss eine TFBS-Tabelle für den Import ausgewählt werden. Die TFBS werden familienweise gespeichert, daher ist als Tabelle für den Import die Tabelle mit dem Namen der entsprechenden TF-Familie zu wählen. Dies geschieht im unteren Bereich des Importfensters. Die obere Tabelle zeigt an, für welche Faktorfamilien bereits TFBS-Tabellen existieren (Abb. 8, „Table“). Die untere Tabelle zeigt an, welche TFBS bereits in der gewählten TFBS-Tabelle vorliegen (Abb. 8, „Content“). Durch Klicken auf den linken grauen Randbereich der oberen Tabelle wird eine TFBS-Tabelle ausgewählt, im Randbereich erscheint dann ein schwarzer Pfeil, und die untere Tabelle wechselt auf den Inhalt der gewählten TFBS-Tabelle.

Für den Import kann in der oberen Tabelle eine TFBS-Tabelle ausgewählt werden, alternativ kann mithilfe des Knopfes „Create new table“ eine neue TFBS-Tabelle

erstellt werden. Wenn die „Create new table“-Funktion benutzt wird, wird der Benutzer nach einem Namen für die neue Tabelle gefragt. Nach Eingabe des Namens und Drücken auf „OK“ wird die neue Tabelle erstellt und angezeigt. Nach Auswahl der Tabelle müssen noch ein Faktornamen sowie eine Beschreibung für die zu importierenden TFBS in die Textfelder („factor name“ und „description“, Abb. 8) im unteren Bereich des Formulars eingegeben werden, anschließend kann der „Import“-Knopf gedrückt werden. Die TFBS aus dem PATSER-Screening werden nun in die Datenbank importiert, der Faktornamen und Beschreibung werden nach dem Import der TFBS in die Content-Tabelle übernommen. Nach Abschluss des Importvorganges wird automatisch die Anzahl der importierten TFBS angezeigt und in die AthaMap Datenbank eingetragen.

Zusätzliche Informationen zur Darstellung im AthaMap Webinterface wie der min. und max. Score, Threshold, Spezies, Matrize, Referenz, Anzahl der detektierten TFBS sowie eventuell ein restriktiver Score für Colokalisationsanalysen müssen manuell mit dem Tool SQL Server Enterprise Manager in die Tabelle „factor\_info“ eingetragen werden.

### 2.5 Identifizierung von kombinatorischen Elementen

Die Identifizierung von kombinatorischen Elementen erfolgte mithilfe der Colokalisationsfunktion von Martin Schindler, die in AthaMap implementiert wurde. Da die öffentlich zugängliche Version der Colokalisationsfunktion eine Distanzbeschränkung auf maximal 50bp enthält, wurde eine Version ohne diese Beschränkung verwendet. Als Browser für die Durchführung von Colokalisationsanalysen wurde der Microsoft Internet Explorer benutzt, da dies die Weiterverarbeitung des Ergebnisses vereinfacht. Zur Detektion potenzieller kombinatorischer Elemente wurden die Elemente und Parameter verwendet, die nachfolgend aufgeführt sind (Tab. 2).

Tabelle 2: Parameter zur Detektion potenzieller kombinatorischer Elemente

Kombinatorisches Element	Distanz zwischen Einzelementen	Verwendete TFBS
bZIP/bZIP	10-15bp	TGA1a/TGA1a
bZIP/DOF	3-40bp	TGA1a/DOF
MYC/MYB	3-40bp	TGA1a/allMYB
MADS/MADS	10-100bp	allMADS/allMADS
RAV	3-20bp	RAV1(1)/RAV1(2)

## 2. Material und Methoden

Für die Detektion der potenziellen kombinatorischen Elemente wurden die jeweils aufgeführten Einzelelemente in den Auswahlfenstern des AthaMap-Interfaces ausgewählt (Abb. 9, „TGA1a“). Die minimale und maximale Distanz (Tab. 2) wurde in die Fenster für „min. Spacer length“ und „max. Spacer length“ eingetragen und die Analyse durch Druck auf den „Search“-Knopf gestartet.

**AthaMap - Microsoft Internet Explorer**

Adresse: [http://www.athamap.de/work-403H/search\\_colocalization.php?check=1&factor1=TGA1a&threshold1=5.21&factor2](http://www.athamap.de/work-403H/search_colocalization.php?check=1&factor1=TGA1a&threshold1=5.21&factor2)

# AthaMap

- Home
- Search
- Colocalization**
- Description
- Documentation
- Contact
- Links
- Disclaimer

## Colocalization-Analysis

Transcription-Factor 1:  Min. Threshold:  (5.21 - 10.45)

Transcription-Factor 2:  Min. Threshold:  (5.21 - 10.45)

[Table of restriction scores](#)

Size of colocalization window:

Min. Spacer length:  Max. Spacer length:

Sort By:

**Total number of combinatorial elements detected: 3438**

Transcription-Factor 1:

Name: TGA1a  
Species: Nicotiana tabacum  
Family: bZIP  
Min. Threshold: 5.21  
Number TFBSs: - in database: 142072  
- used in this analysis: 142072

Transcription-Factor 2:

Name: TGA1a  
Species: Nicotiana tabacum  
Family: bZIP  
Min. Threshold: 5.21  
Number TFBSs: - in database: 142072  
- used in this analysis: 142072

Size of colocalization window:  
Spacer length: 10-15

[Show overview](#)

Position		Chromosome	Orientation		Spacer
Factor 1	Factor 2		Factor 1	Factor 2	
83354	83343	1	-->	-->	11
83354	83343	1	-->	<--	11
83354	83343	1	<--	-->	11
83354	83343	1	<--	<--	11
84492	84479	1	<--	-->	13
84492	84479	1	-->	-->	13

Abb. 9: Das AthaMap-Webinterface zur Durchführung von Colokalisationsanalysen

Das tabellarische Suchergebnis von AthaMap wurde mittels Microsoft Excel in eine Textdatei umgewandelt und abgespeichert wie nachfolgend beschrieben.

Die Tabelle mit den einzelnen Positionen von Colokalisationen (Abb. 9) wird kopiert und in ein Excel-Arbeitsblatt eingefügt. Beim Einfügen wird automatisch der Textimport-Assistent von Excel aktiviert und fragt nach der Formatierung beim Import; hier ist die Formatierung der Zielzellen zu übernehmen, ansonsten werden die HTML-Links aus der ursprünglichen Formatierung übernommen.

## 2. Material und Methoden

Sobald die Tabelle in Excel eingefügt ist, wird sie im Format „Text; Tabstop getrennt“ abgespeichert, um sie anschließend wie weiter unten beschrieben mit dem Importprogramm „CEImporter“ (Abschnitt 2.6) in die AthaMap Datenbank zu importieren.

### 2.6 Import von kombinatorischen Elementen

Für den Import von kombinatorischen Elementen, die mithilfe der AthaMap Colokalisationsfunktion identifiziert wurden, wurde das Tool „CEImporter“ entwickelt (Abb. 10). CEImporter wurde in Visual Basic 6.0 entwickelt und läuft auf jedem Windows-PC ab Win95. Zum Import von Suchergebnissen der AthaMap-Colokalisationsfunktion müssen diese zunächst wie unter Abschnitt 2.5 beschrieben in ein Tabstop-getrenntes Textformat umgewandelt werden.

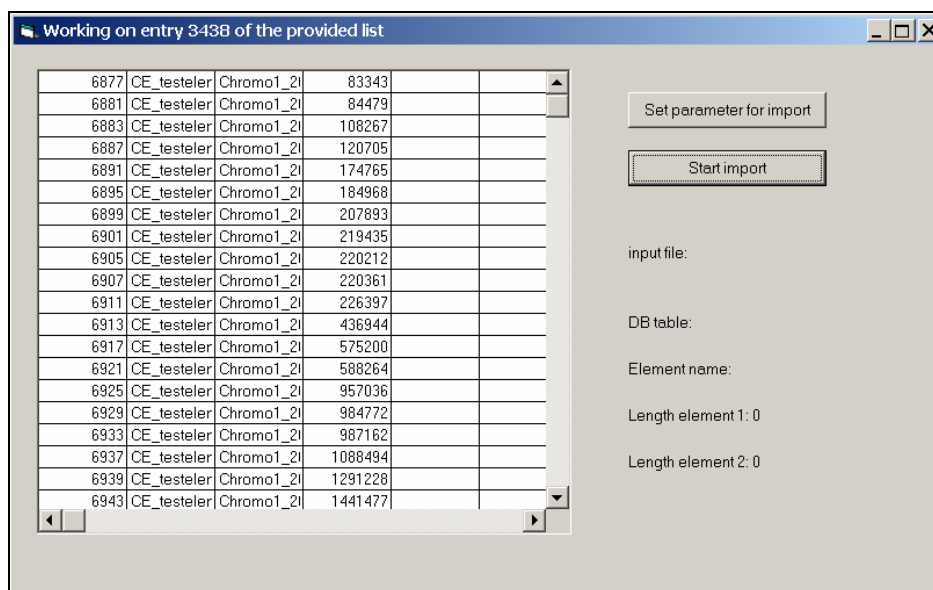


Abb. 10: Das Tool CEImporter für den Import von kombinatorischen Elementen, die mithilfe der Colokalisationsfunktion identifiziert wurden.

Nach dem Start des Tools müssen zunächst die Parameter für den Import festgelegt werden. Hierzu wird der Knopf „Set parameter for import“ gedrückt (Abb. 10), daraufhin erscheinen mehrere Dialogfenster, in denen diese Parameter abgefragt werden (nicht gezeigt). Zunächst wird der Benutzer aufgefordert, eine Datei für den Import zu wählen. Hier ist die Textdatei mit den Ergebnissen einer Colokalisationsanalyse zu wählen, die wie unter Abschnitt 2.5 beschrieben erstellt wurde.

Im nächsten Dialogfenster muss der Name der Datenbank-Tabelle für den Import eingegeben werden. Diese Tabelle wird genauso wie eine TFBS-Tabelle vor dem Import mit den TFBSitesTools (Abschnitt 2.4, Abb. 8, „Create new table“) generiert. Als nächstes wird ein Name für das zu importierende Element abgefragt. Zu guter letzt muss die Länge der Matrize von Element 1 und Element 2 der Colokalisationsanalyse eingegeben werden.

Die vom Benutzer eingegebenen Parameter werden gespeichert und im Programmfeld von CEImporter angezeigt. Nach Druck auf den „Start import“-Knopf (Abb. 10) beginnt das Programm mit dem Import der kombinatorischen Elemente aus der angegebenen Textdatei in die gewünschte Tabelle. Die Anzahl der bereits abgearbeiteten Einträge wird in der Kopfzeile des Programms angezeigt (Abb. 10). Nach dem Import müssen zur Anzeige in AthaMap noch die Daten des kombinatorischen Elementes in die AthaMap-Tabelle „combinatorial\_info“ eingetragen werden. Hier sind der Name des Elementes einzutragen, die tolerierte Distanz zwischen den Einzelelementen, die verwendeten Matrizen und die Anzahl der detektierten kombinatorischen Elemente.

### 2.7 Allgemeiner Aufbau der AthaMap-Datenbank

Als Basis für den Betrieb der AthaMap Datenbank dient ein Microsoft-SQL-Serversystem, das unter dem Betriebssystem Windows 2000 Server Edition betrieben wird. Die Hardware des Servers beinhaltet ein Dual-Prozessorsystem mit zwei Pentium III 1000Mhz-CPU's mit 2 Gigabyte RAM und einem RAID Level 5 SCSI-Festplatten-System.

Die AthaMap Datenbank besteht aus mehreren miteinander verbundener Tabellen (Abb. 11), die sich im Prinzip in drei Typen einteilen lassen; 1. Tabellen zur Speicherung der TFBS-Daten (Position, Orientierung, Score); 2. Tabellen mit Annotationsdaten des *Arabidopsis*-Genoms, die z.B. zur Visualisierung benötigt werden oder um die Lokalisierung eines Elementes zu bestimmen; 3. zusätzliche Tabellen mit Informationen zu den einzelnen TFs, kombinatorischen Elementen etc. Die TFBS werden in den TFBS-Tabellen gespeichert, die normalerweise den Namen der Faktorfamilie tragen (Abb. 11, bZIP und MADS). Eine Ausnahme bildet hier die Tabelle namens „other“, in der TFBS von TFs gespeichert sind, von denen nur jeweils ein Mitglied der entsprechenden TF-Familie in AthaMap aufgenommen wurde

(nicht gezeigt). Zu jeder TFBS-Tabelle existiert eine cat-Tabelle, die das Inhaltsverzeichnis der TFBS-Tabelle darstellt. Sie gibt an, zu welchen TFs die TFBS in der zugehörigen TFBS-Tabelle gehören und wie viele Bindungsstellen für jeden TF gespeichert wurden. Die TFBS-Tabellen und die zugehörigen cat-Tabellen werden immer gemeinsam mithilfe der TFBSitesTools (Abschnitt 2.4) erzeugt.

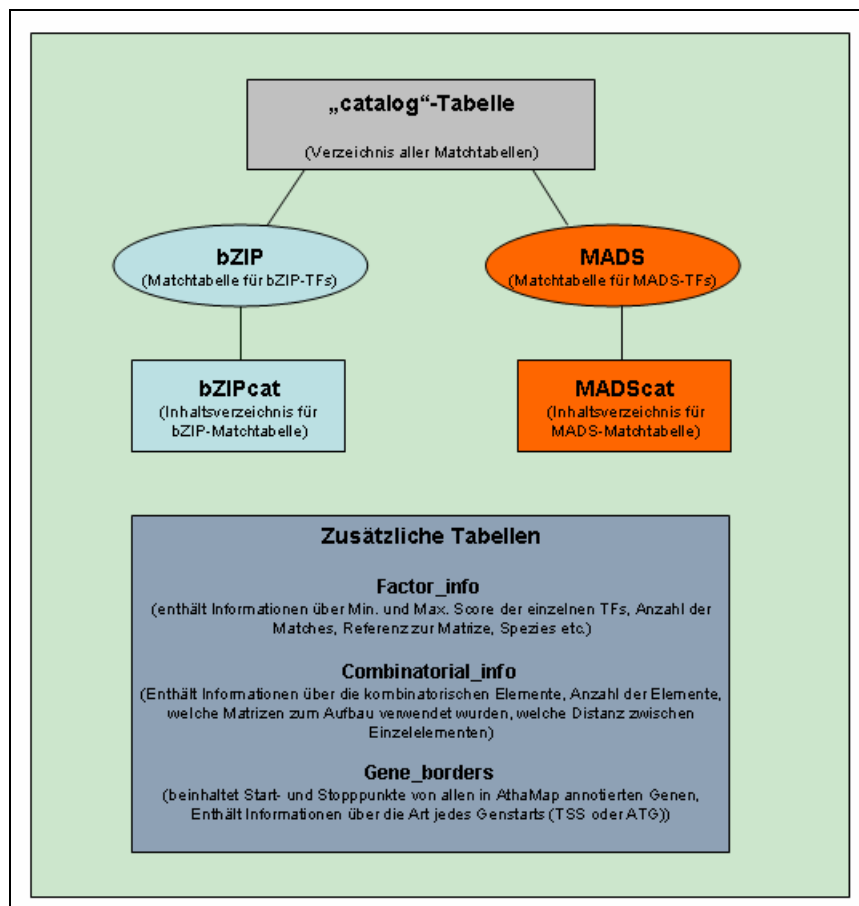


Abb. 11: Allgemeiner Aufbau der AthaMap Datenbank.

Die „Catalog“-Tabelle dient als Verzeichnis, in dem die Namen aller anderen TFBS-Tabellen verzeichnet sind.

Außerdem werden Tabellen benötigt, die zusätzliche Informationen für die Darstellung im AthaMap Webinterface enthalten. In der Tabelle „gene\_borders“ stehen Daten über Anfangsposition und Endposition jedes Gens sowie seine Orientierung. In der Tabelle „Factor\_info“ sind die in AthaMap angezeigten Daten für die einzelne TFs gespeichert. Die Tabelle „combinatorial\_info“ enthält die Informationen über kombinatorische Elemente, die auf der „Documentation“ Seite angezeigt werden.

## 2.8 Bestimmung der Position von TFBS in AthaMap relativ zum ATG

Die Positionsanalyse von in AthaMap gespeicherten TFBS wurde mit dem Werkzeug TFBSitesTools durchgeführt (Abschnitt 2.4, Abb. 6). Anstelle des „Import“-Knopfes drückt man nach dem Programmstart den Knopf „Analyze one factor“. Dadurch wird eine Auswahloberfläche sichtbar (Abb. 12), auf der eine TF-Familie und anschließend ein TF ausgewählt werden kann.

analyse

Selected table: bZIP; Selected factor: 1 TGA1; Description: genomic screening; auto threshold  
Number of records total: 53494 selected: 17177 Lower threshold=8  
Exclude redundant records tagged ++4; ++4;

Close

Select factor Analyse Core region / CDS Statistics Copy Position count

Please select a table and one factor for analysis.

Table

table_no	table_name	creation
1	bZIP	24.07.2004 10:00:00
2	AP2	24.07.2004 11:00:00
3	homeoZIP	24.07.2004 12:00:00
4	MADS	24.07.2004 13:00:00
5	MYB	24.07.2004 14:00:00
6	other	24.07.2004 15:00:00
7	TCF	24.07.2004 16:00:00

1/28

Content

import_no	factor_name	description	number	c
1	TGA1	genomic screening; auto th	53494	2
4	bZIP910(2)	genomic screening; auto th	470	2
5	bZIP911(1)	genomic screening; auto th	123	2
6	bZIP911(2)	genomic screening; auto th	145	2
8	ABF1	genomic screening; auto th	2419	2

1/8

Threshold

8 Go

☐ only relative positions from -2000  
☒ exclude redundant records -1  
☐ exclude non-core sequence

factor_no	factor_name	chromosome	position	s
8	TGA1	Chromo1_2004.txt	29881	+
12	TGA1	Chromo1_2004.txt	39416	+
13	TGA1	Chromo1_2004.txt	40277	+
22	TGA1	Chromo1_2004.txt	52633	-
28	TGA1	Chromo1_2004.txt	67388	+

1/17177

Abb. 12: Auswahlfenster in TFBSitesTools zur Auswahl eines TFs für die Lokalisierungsanalyse relativ zum nächsten Start-ATG. Die Auswahl der Tabelle für die TF-Familie und des TFs erfolgt analog wie unter Abschnitt 2.4 beschrieben.

Nach der Auswahl wird durch Druck auf die Lasche „Core region / CDS“ die Analysefunktion für die Lokalisierung von TFBS relativ zum ATG in den Vordergrund gebracht (Abb. 13, nächste Seite).

Anschließend wird durch Druck auf den Knopf „CDS“ die Analyse gestartet. Das Programm geht sämtliche TFBS durch und schreibt die relative Position zum nächsten Start-ATG in die Tabellenspalte „gene\_position“ (nicht gezeigt).

## 2. Material und Methoden

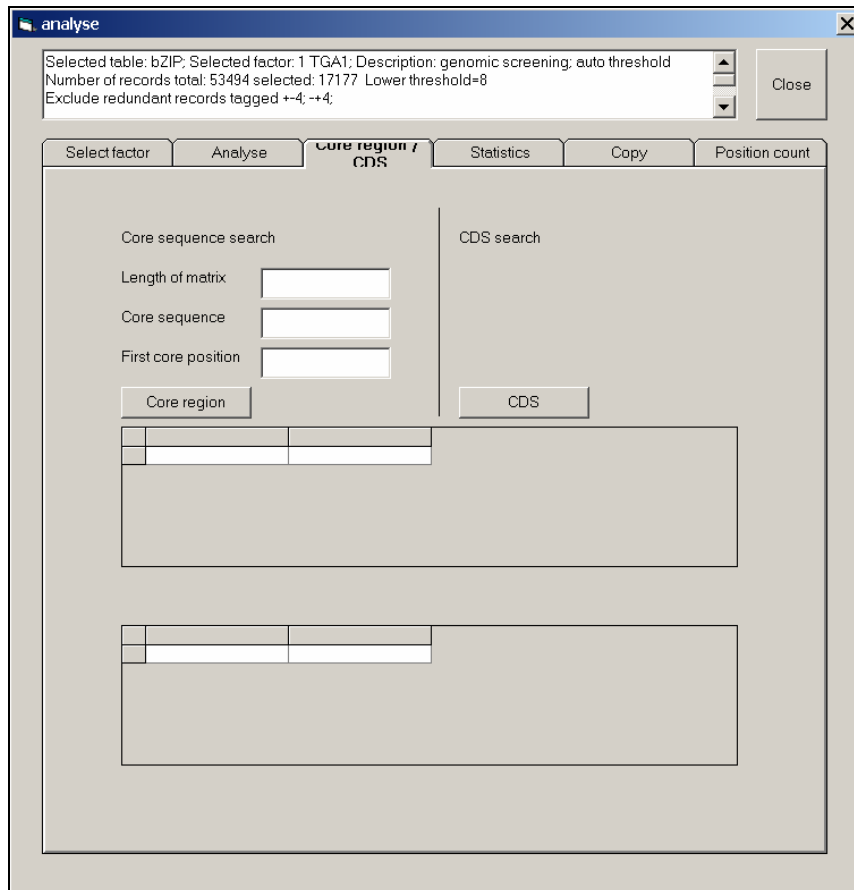


Abb. 13: Analysefunktion für die Annotation der Position jeder TFBS relativ zum nächsten Start-ATG.

### 2.9 Das Tool CE Cleaner zur Entfernung von überlappenden kombinatorischen Elementen in AthaMap

Im Falle der kombinatorischen Elemente, die aus mindestens einer palindromischen TFBS aufgebaut sind, kann es zu sich überlagernden Elementen kommen, die sich nur in der Orientierung der beteiligten Bindungsstellen unterscheiden, von ihrer Position aber identisch sind. Zur Markierung dieser doppelten Elemente wurde das Tool „CE Cleaner“ entwickelt, das Elemente gemäß den einstellbaren Vorgaben für die spätere Entfernung aus der Datenbank markiert.



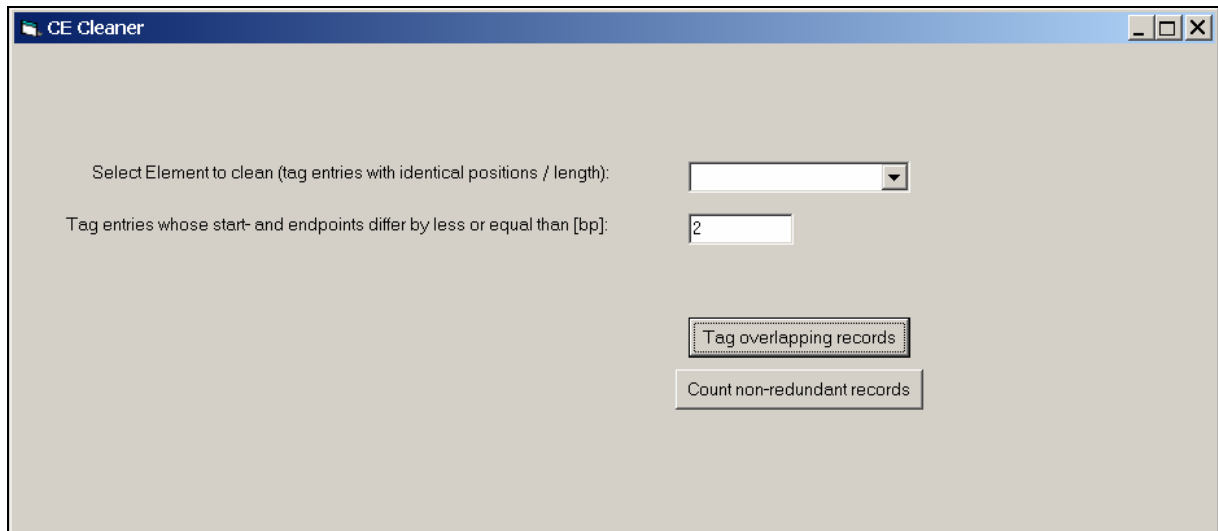


Abb. 14: Das Tool CE Cleaner zur Entfernung überlappender kombinatorischer Elemente

Beim Programmstart werden im oberen Auswahlfenster (Abb. 14) alle Elemente aus der AthaMap-Tabelle „combinatorial info“ (Abb. 11) automatisch zur Auswahl angeboten. Zunächst wählt der Benutzer in diesem Auswahlfenster ein kombinatorisches Element aus, anschließend wird im Textfenster darunter das Ausschlusskriterium für überlappende Elemente definiert. Nach Druck auf den „Tag overlapping...“-Knopf geht das Programm sämtliche Einträge des gewählten Elements durch und prüft paarweise aufeinanderfolgende Einträge auf Überlappung. Wird ein überlappendes Element gefunden, so wird der Eintrag in der Spalte „Sequence“ der TFBS-Tabelle mit dem Wort „doppelt“ markiert (nicht gezeigt). Die Anzahl bearbeiteter Einträge erscheint in der Kopfzeile des Programms CECleaner. Nach Markierung der überlappenden Einträge kann die Anzahl der verbleibenden Einträge durch Druck auf den Knopf „Count non-redundant records“ angezeigt werden.

Da die Überprüfung paarweise erfolgt, wird ein drittes Element, das mit zwei Elementen davor überlappt, nicht markiert. Elemente, die bereits als überlappend markiert sind, werden aber nicht erneut in Analysen mit einbezogen. Also können nach jedem Programmdurchlauf andere Elemente in der Datenbank benachbart sein, die nun ebenfalls überlappen könnten. Aus diesem Grund muss die Funktion „Tag overlapping records“ so oft nacheinander durchgeführt werden, bis sich die Anzahl nicht redundanter Einträge nach dem Funktionsdurchlauf nicht mehr ändert.

Abschließend können die markierten Einträge mithilfe des SQL Server Enterprise Managers dauerhaft entfernt werden. Zu diesem Zweck wechselt man in die Ansicht der entsprechenden Tabelle und gibt folgenden SQL-Befehl ein:

„DELETE from Tabellenname WHERE SEQUENCE = 'doppelt' “ Durch Ausführung des Befehls werden alle markierten Einträge gelöscht.

### **2.10 Bestimmung der relativen Position zur Genstruktur von Einzel- und kombinatorischen Elementen**

Um festzustellen, in welcher Annotationseinheit (Intron, Exon, UTR) sich eine TFBS oder ein kombinatorisches Element befindet, wurde das Tool „Localizer“ benutzt (Abb. 15, nächste Seite). Als Eingabe wird vom Benutzer eine Textdatei im Format „Tabstop-getrennt“ verlangt, die Positionsangaben und dahinter Chromosomenbezeichner in Form von Nummern (1-5) enthält. Diese Datei erhält man aus dem HTML-Output der AthaMap-Colokalisationsfunktion durch Verarbeitung mit Excel wie im Folgenden beschrieben.

Es wird mit dem Internet-Explorer eine Colokalisationsanalyse in AthaMap durchgeführt (Abschnitt 2.5). Der tabellarische Output wird ohne die Spaltenüberschriften kopiert und in MS Excel in eine leere Tabelle eingefügt. Dabei werden die einzelnen Spalten der Tabelle automatisch getrennt. Von dieser neuen Excel-Tabelle wird die erste Spalte mit den Positionen des als erstes gewählten Elementes kopiert und in eine zweite leere Tabelle eingefügt. Auf dieselbe Weise wird die Spalte mit den Chromosomennummern kopiert und ebenfalls in die zweite Tabelle eingefügt. Es liegen in dieser neuen Tabelle nun zwei Spalten vor, die erste enthält die Positionen des für die Colokalisationsanalyse als erstes gewählten Elementes, die zweite Spalte die zugehörigen Chromosomenbezeichner. Diese zweite Tabelle wird nun aus Excel im Textformat abgespeichert („Speichern unter...“, Dateityp: „Text Tabstop getrennt“). Sie dient als Eingabedatei für die Positionen der Elemente, für die nachfolgend die Annotationseinheit bestimmt werden soll.

Der Localizer fragt nach dem Druck auf den „Localize!“-Knopf nach einer Eingabedatei und nach einem Ort zum Speichern der Ausgabedatei (Abb. 15, nächste Seite). Als Eingabedatei wird die Excel-Textdatei, die nach der Beschreibung oben erstellt wurde, ausgewählt. Nach Eingabe eines Namens für die Ausgabedatei wird diese automatisch im Textformat erstellt, sie enthält eine Zusammenfassung der eingespeisten Positionsangaben, und in welchen Annotationseinheiten sie lokalisieren.

## 2. Material und Methoden

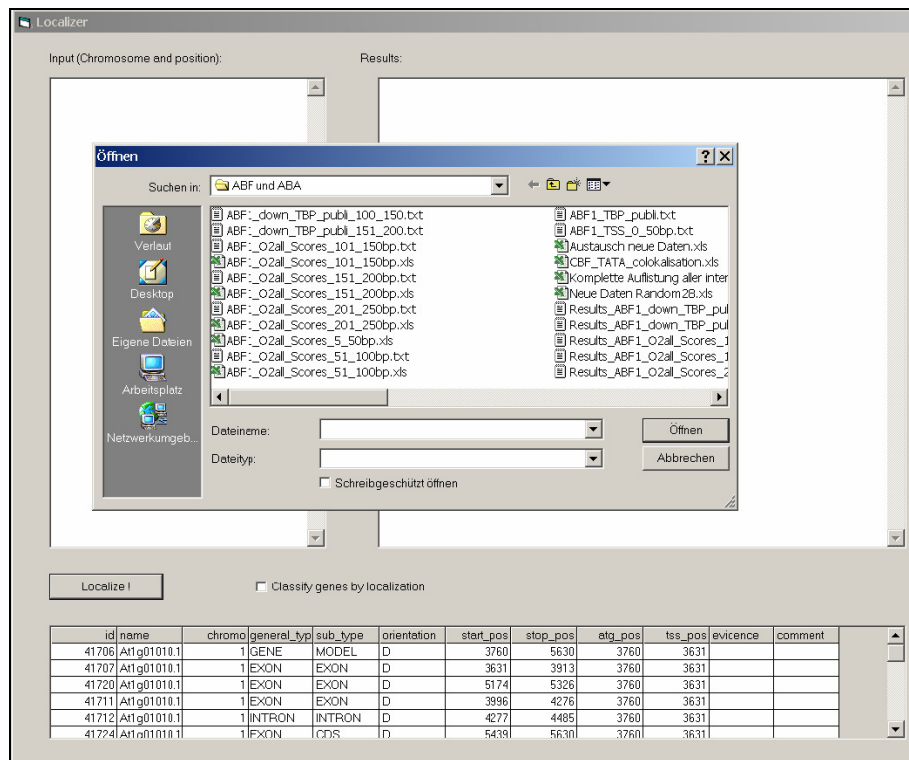


Abb 15: Screenshot der Localizer-Bedienoberfläche mit Dialogfenster zur Dateiauswahl

Hierbei wird zwischen Exon-, Intron- und intergenischer Lokalisierung unterschieden. Zunächst erfolgt eine Ausgabe für jede einzelne eingespeiste Position, anschließend werden die AGIs der jeweiligen Gene nochmals gruppenweise sortiert nach Exon, Intron oder intergenischer Lokalisierung ausgegeben. Die Ausgabe des Ergebnisses erfolgt sowohl in einem Textfenster des Programms als auch in der Ausgabedatei. Diese AGIs können nachfolgend zur weiteren Analyse in andere Tools eingespeist werden. In der vorliegenden Arbeit wurde das Localizer-Tool dazu verwendet, um intergenische Colokalisationen zu identifizieren und um anschließend den Expressionsstatus der entsprechenden Gene aus der Genevestigator-Datenbank abzufragen (Abschnitt 2.15).

Die präzise Bestimmung der Annotationseinheit erfolgt mithilfe der Annotationstabelle „gene\_data“, die die genauen Grenzen der einzelnen Annotationseinheiten enthält (Abb. 15, unten). Die Informationen in dieser Tabelle wurden durch Heraustrennen mit einem PERL-Skript aus der TIGR-Annotationsdatei extrahiert. Die „gene\_data“-Tabelle wurde mir freundlicherweise von Claudia Galuschka zur Verfügung gestellt (persönliche Mitteilung).

### 2.11 Positionell eingeschränkte Annotation von TBP- und CBF-Bindungsstellen in AthaMap2004

Um potenzielle Bindungsstellen für die Faktoren TBP und CBF in AthaMap positionell eingeschränkt zu annotieren sowie für die Deletion von positionell identischen Bindungsstellen aus der Datenbank, wurde das Werkzeug „TBPCurator“ entwickelt (Abb. 16).

Für die Annotation wurde zunächst mit beiden Matrizen gescreent und die Ergebnisse ungefiltert in die AthaMap Datenbank importiert (Tabellen TBP\_full und CCAAT\_full).

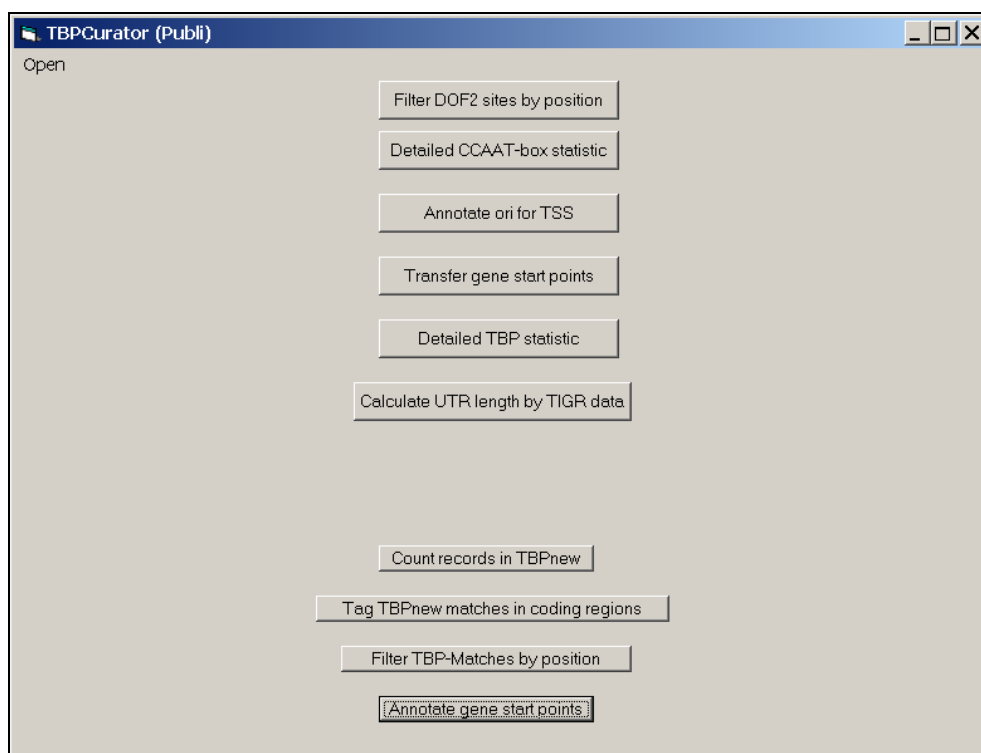


Abb. 16: Die Bedienoberfläche von TBPCurator

Der Knopf „Filter TBP-Matches by position“ (Abb. 16) dient zur positionellen Einschränkung der TATA-Boxen. Ausgehend von der Tabelle „gene\_borders“ (siehe Abschnitt 2.7), in der die Start- und Stoppunkte aller proteincodierenden Gene verzeichnet sind, wird für jedes Gen einzeln überprüft, ob sich im upstream-Bereich des Gens TBP- oder CBF-Bindungsstellen befinden oder nicht. Soweit diese vorhanden sind, werden sie extrahiert und in die neue TBP-Tabelle übertragen. Die Größe der Annotationsbereiche wurde aufgrund eigener Analysen festgelegt. Die Parameter für die Größe der Annotationsbereiche sind im Programmcode von TBPCurator definiert und können nur dort verändert werden.

Mithilfe von TBPCurator wurden die TFBS vom TBP und des CBF positionell eingeschränkt annotiert.

### 2.12 Entfernung von positionsgleichen TBP-Bindungsstellen

Bindungsstellen für das TBP-Protein treten aufgrund ihres palindromischen Charakters häufig an derselben Position, aber mit unterschiedlicher Orientierung auf. Um positionsgleiche TFBS zu vermeiden, werden diese mit dem Tool TBPClean entfernt.

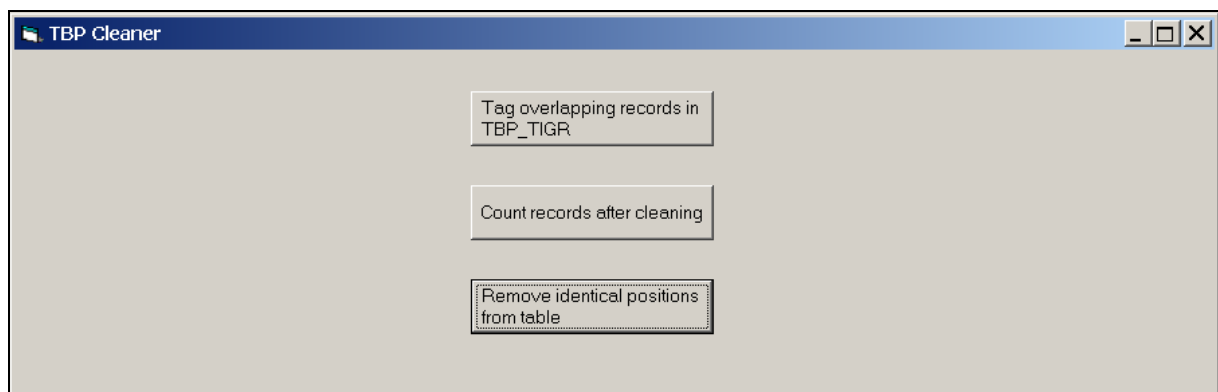


Abb. 17: Das Tool TBPClean zur Entfernung positionsgleicher TFBS des TBPs.

Durch Druck auf den Knopf „Remove identical positions from table“ (Abb. 17) werden diejenigen Einträge in der TBP-Tabelle markiert, die an genau derselben Position liegen wie der vorangegangene Eintrag. Beim Funktionsdurchlauf werden die Einträge immer paarweise überprüft. Sobald ein Eintrag dieselbe Position aufweist wie der Eintrag davor, wird der zweite Eintrag markiert. In dem hypothetischen Fall, dass drei positionsgleiche Einträge hintereinander liegen, würde beim ersten Ausführen der Funktion lediglich der zweite Eintrag markiert werden. Beim erneuten Ausführen der Funktion werden dann der erste und der dritte Eintrag benachbart liegen, und der dritte Eintrag wird markiert. Um alle positionsgleichen Einträge zu entfernen, muss die Funktion „Remove identical...“ also solange ausgeführt werden, bis sich die Gesamtzahl der unmarkierten Einträge nicht mehr ändert. Dies kann mithilfe der Funktion „Count records after cleaning“ festgestellt werden; diese Funktion zeigt nach Druck auf den entsprechenden Knopf die Anzahl unmarkierter Einträge in einem Nachrichtenfenster an.

### 2.13 Bestimmung der durchschnittlichen Länge des 5'UTRs inkl. Introns

Die Funktion „Calculate UTR length by TIGR data“ des Tools „TBP\_Curator“ (Abschnitt 2.11, Abb. 16) diente zur Bestimmung der durchschnittlichen Länge des 5' UTR inklusive aller in diesem Bereich auftretenden Introns. Da 5'UTRs normalerweise keine Introns enthalten, wird darauf hingewiesen, dass im Folgenden der Begriff 5'UTR immer die in diesem Bereich auftretenden Introns mit einschließt. Als Datengrundlage benötigt die Funktion zwei Tabellen in der AthaMap Datenbank; einerseits wird die Tabelle „gene\_models“ benötigt, in der die Annotationsdaten aller in *Arabidopsis* annotierten Gene gespeichert vorliegen. Diese Tabelle wurde aus den öffentlich zugänglichen Annotationsdaten bei TIGR durch Verarbeitung mit einem PERL-Skript hergestellt (Claudia Galuschka, persönliche Mitteilung). Für jedes Gen sind hier seine Orientierung, Start- und Endposition sowie die Position des ATG und, soweit bekannt, die Position des TSS annotiert. Die potenziellen alternativen Transkripte eines Gens sind durch Zahlen kenntlich gemacht; normalerweise besitzt jedes Gen wenigstens ein Transkript, welches in der TIGR-Annotation durch die entsprechende AGI-Nummer und ein angehängtes „.1“ bezeichnet wird. Im Falle des Auftretens von alternativen Transkripten eines Gens sind diese durch fortlaufende Nummern (.2 etc) gekennzeichnet. Für die Berechnung der durchschnittlichen 5'UTR-Länge wurden, wenn möglich, die .1-Transkripte herangezogen. Für die Speicherung der 5'UTR-Längen der Gene wird eine Tabelle mit Namen „UTR\_TIGR“ benötigt, die außer einem Primärschlüssel die Attribute „gene\_name“ und „UTR\_length“ enthält. Diese Tabelle wird manuell mit dem SQL Server Enterprise Manager erstellt. Bei der Bestimmung der 5'UTR-Länge geht die Programmfunktion nacheinander alle Gene mit annotiertem TSS durch und fragt die jeweilige Position von ATG und TSS ab. Die Funktion berechnet aus diesen Daten die Länge des 5'UTRs und schreibt den Namen des zugehörigen Gens sowie die 5'UTR-Länge in die Tabellenspalte der UTR\_TIGR-Tabelle. Aus der UTR\_TIGR-Tabelle können die Daten der Primärtranskripte (also der .1-Transkripte) mit einem SQL-Befehl abgefragt werden. Um den SQL-Befehl einzugeben, öffnet man auf einem Windows-PC das Programm SQL Server Enterprise Manager, klickt auf die Datenbank AthaMapWork und wählt den Unterpunkt „Tabellen“. Anschließend macht man einen Rechtsklick auf die Tabelle UTR\_TIGR und wählt im Kontextmenü den Befehl „Tabelle öffnen“. Es erscheint eine Passwortabfrage, nach Eingabe des Passwortes werden die ersten

Zeilen der Tabelle angezeigt. Durch Druck auf den Knopf mit der Beschriftung „SQL“ wird der SQL-Bereich eingeblendet, in den man den folgenden SQL-Befehl eingeben kann:

„Select \* from UTR\_TIGR where gene\_name like ‚%.1‘ order by gene\_name“.

Nach Druck auf den Knopf mit dem Ausrufungszeichen wird der Befehl ausgeführt, und die UTR-Längen der 16315 Primärtranskripte mit annotiertem TSS werden angezeigt. Die Tabelle wird aus dem SQL Server Enterprise Manager durch Kopieren und Einfügen in Excel übernommen und dort weiter verarbeitet.

Zusätzlich zu den 16315 Primärtranskripten gibt es noch vier weitere Ausnahmen von Genen mit annotiertem TSS, für die jedoch kein .1-Transkript annotiert ist, sondern nur ein .2-Transkript (namentlich: At1g45201.2; At1g45249.2; At1g77080.2; At5g21020.2.) Von diesen Genen müssen die Daten manuell aus der Tabelle UTR\_TIGR abgefragt werden. (Beispiel: „Select \* from UTR\_TIGR where gene\_name = 'At1g45201.2'“) Auch die Daten dieser vier Gene werden in die Excel-Tabelle übernommen, so dass nun die Daten aller 16319 Gene mit annotiertem TSS vorliegen. Nun wird von den Werten in der Tabellenspalte mit der UTR-Länge mithilfe der „Mittelwert“-Funktion von Excel der Mittelwert bestimmt.

### **2.14 Selektion zufällig gewählter Gene mittels „RandomGeneSelector“**

Um die Signifikanz des Auftretens von colokalisierenden Elementen in einer Gruppe von Genen mit einem Hintergrundmodell vergleichen zu können, wurden Gruppen mit derselben Anzahl zufällig ausgewählter Gene überprüft. Die zufällige Auswahl der Gene erfolgte mit dem Visual Basic Werkzeug „RandomGeneSelector“ (Abb. 18, nächste Seite).

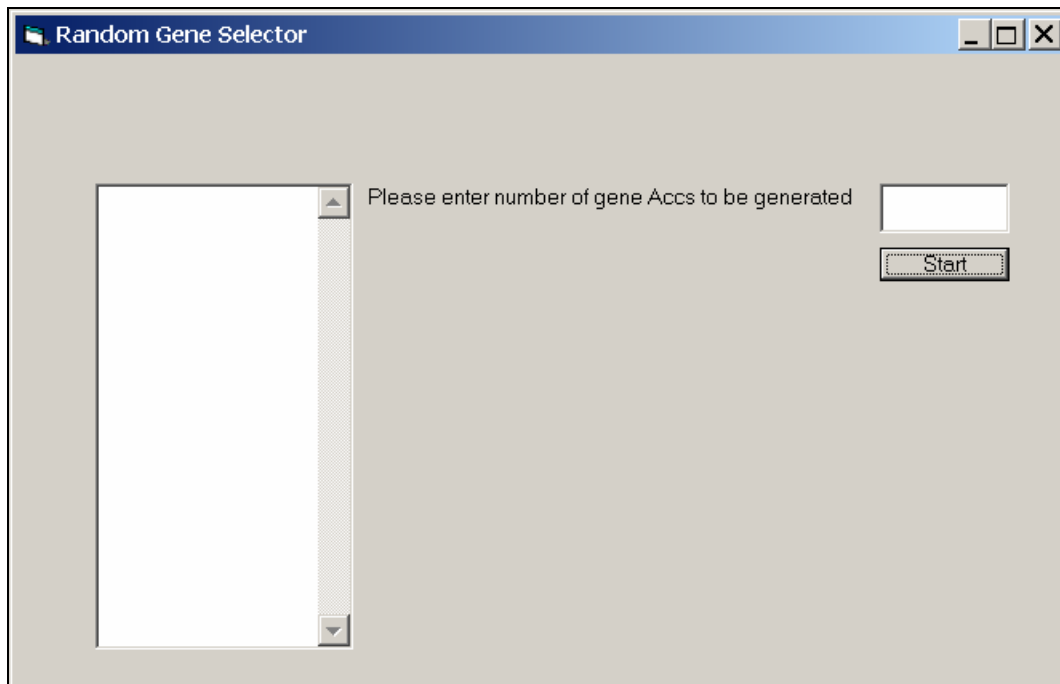


Abb. 18: Bedienoberfläche des Programms RandomGeneSelector

Nach dem Start des Programms gibt der Benutzer die gewünschte Anzahl zufällig zu erzeugender Gene in das Textfenster über dem Start-Knopf ein.

Nach Druck auf den Startknopf wird zunächst die eingebaute RND-Funktion (Zufallsgenerator) von Visual Basic mit der momentanen Windows-Uhrzeit initialisiert, dies stellt sicher dass nicht zweimal dieselbe „zufällige“ Reihe erzeugt wird.

Anschließend wird die in Visual Basic eingebaute RND-Funktion gestartet und erzeugt eine Zufallszahl zwischen 1 und 29993. Diese Zahl entspricht der Nummer eines der 29993 annotierten Gene in der AthaMap-Annotationstabelle. Durch eine Datenbankabfrage aus der Annotationstabelle wird bestimmt, welche AGI das betreffende Gen besitzt, und diese wird im Ergebnistextfenster auf der linken Seite ausgegeben. Dieser Zyklus wiederholt sich, bis die gewünschte Anzahl zufällig zu wählender Gene erzeugt ist. Es erfolgt keine automatische Kontrolle auf doppelt auftretende Gene in der Zufallsauswahl, da die zufällige zweimalige Auswahl eines Gens bei fast 30.000 Genen relativ unwahrscheinlich ist. Das doppelte Auftreten eines Gens in einer der Zufallskontrollgruppen wurde aber durch manuelle Kontrolle ausgeschlossen.



## 2.15 Korrelation von *in silico* Daten mit Microarray-basierten Genexpressionsprofilen

Um das Auftreten von Colokalisationen mit Genexpressionsprofilen korrelieren zu können, wurde auf die Genevestigator-Datenbank ([www.genevestigator.ethz.ch](http://www.genevestigator.ethz.ch)) zurückgegriffen, die Microarray-basierte Genexpressionsprofile öffentlich zugänglich im Internet zur Verfügung stellt [Zimmermann et al., 2004].

**Meta-Analyzer**

**Introduction**  
The Meta-Analyzer was created to analyze gene families or other lists of genes. It queries the database to answer the following questions:  
 • What are the expression profiles of genes from a given list relative to organ, response or growth stage?  
 • Which gene profiles cluster together?

**Submission form**

**Selected array type**  
 ATH1: 22k array ☐ Wild-type

**Selected ID's (max 50)**  
 261585\_at|AT1G01010

**Hierarchical Clustering**  
 Distance: Euclidean Linkage: Average linkage

**Meta Analysis by:**  
 Plant Organs ☐ Exclude List:

**Scale Type (Data only)**  
☒ Linear ☐ Log2(n)

**Result**  
 Show Color Legend

Probeset	0 callus	1 cell suspension	2 seedling	21 cotyledons	22 hypocotyl	23 radicle	3 inflorescence	31 flower	311 ovary	3112 stigma	312 petal	313 sepal	314 stamen	315 pedicel	32 silique	33 seed	34 stem	35 node	36 shoot apex	37 cauline leaf	4 rosette	41 juvenile leaf	42 adult leaf	43 petiole	44 senescent leaf	5 roots	52 lateral root	55 elongation zone
261585_at	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575
AT1G01010	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575	575

Abb. 19: Das „Meta Analyzer“-Webinterface der Genevestigator Datenbank.

Nach Eingabe der Benutzerdaten wurde als Chiptyp der „22k“-Chip von Affymetrix ausgewählt (Abb. 19: ATH1). Daten über Lokalisation wurden mit dem Localizer-Tool (Abschnitt 2.10) erstellt, und die AGIs derjenigen Gene, die intergenische Colokalisationen aufweisen, in Genevestigator weiter analysiert. Nach Drücken des „Add“-Buttons im Metaanalyzer Webinterface (Abb. 19) und Eingabe der AGI-Nummern aus der Localizer Ausgabedatei wird zunächst durch Genevestigator überprüft, welche der eingespeisten Gene auf dem Chip vorhanden sind. In Fällen, wo ein Gen nicht als Sonde („probe“) auf dem gewählten Chiptyp repräsentiert ist, erhält der Benutzer für das entsprechende Gen eine Fehlermeldung. In einigen Fällen ist für ein Gen mehr als eine Sonde verfügbar, in diesem Fall muss der Benutzer eine oder mehrere Sonden auswählen, die dann für die Abfrage des Expressionsstatus des entsprechenden Gens verwendet wird. In der vorliegenden

Untersuchung wurde in den Fällen, wo mehrere Sonden für ein Gen verfügbar waren, die jeweils erste Sonde ausgewählt.

Nach Auswahl der Sonden und Bestätigung durch den „OK“-Button werden die eingegebenen AGIs im Genevestigator-Fenster unter „Selected ID's“ dargestellt. Die Parameter „Distance“ und „Linkage“ wurden unverändert bei den voreingestellten Werten belassen („Euclidian“ und „Average linkage“). Um an die stressinduzierten Genexpressionsprofile zu gelangen wurde „Meta analysis by Stress Responses“ gewählt. Anschließend wurden die Genexpressionsprofile der eingespeisten Gene durch Drücken des „Data“-Buttons abgerufen und im Genevestigator-Fenster dargestellt.

Der Expressionsstatus der abgefragten Gene wird als Tabelle dargestellt; signifikante Geninduktion ist durch rote Farbmarkierung der entsprechenden Zelle gekennzeichnet, Repression durch grüne Farbmarkierung. In Fällen, wo die Expression eines Gens nicht signifikant verändert ist, wird ein schwarzes Feld angezeigt. Zusätzlich ist der Faktor der Induktion bzw. Repression in dem entsprechenden Feld als Zahlenwert aufgeführt.

### **2.16 Implementierung des AthaMap Webinterfaces**

Das AthaMap Webinterface ([www.AthaMap.de](http://www.AthaMap.de)) wurde von Martin Schindler in der Sprache HTML entwickelt. Sämtliche Programmfunktionen dieser Oberfläche, wie z.B. Datenbankzugriffsfunktionen, wurden von Martin Schindler in der Sprache PHP implementiert. Die Funktion für die Colokalisationsanalyse („fnFindFactorNear“) wurde von Martin Schindler in der Sprache SQL entwickelt.

### **2.17 Statistische Auswertung mittels t-Test**

Die statistische Auswertung der Colokalisationsanalysen erfolgte mithilfe des t-Testes, wie in Zar „Biostatistical analysis“, Kapitel 8.2, beschrieben [Zar, 1984]. Um festzustellen, ob innerhalb einer Gengruppe („Prädiktionsgruppe“) signifikant viele Gene auftreten, die durch einen bestimmten Stimulus induziert werden, wurde die entsprechende Gengruppe mit fünf zufällig zusammengestellten Kontrollgruppen verglichen.

Zunächst wurde das Verhalten der Gene der Prädiktionsgruppe untersucht und festgehalten, wie viele Gene induziert, reprimiert oder nicht reguliert vorlagen. Anschließend wurden fünf Gruppen gleicher Größe aus zufällig gewählten Genen gebildet und das Verhalten dieser Gene gruppenweise festgehalten. Aus den Werten der fünf Zufallsgruppen wurde ein Mittelwert gebildet, anschließend wurde mittels t-Test überprüft, ob die Anzahl induzierter Gene in der Prädiktionsgruppe signifikant vom Durchschnitt der Anzahl der induzierten Gene in den Zufallskontrollgruppen abweicht.

Folgende Formel wurde verwendet:

$$t = \frac{X - \bar{X}}{\frac{s}{\sqrt{n}}}$$

Abb. 20 Formel für den t-Test

$X$  = Anzahl induzierter Gene in der Prädiktionsgruppe

$\bar{X}$  = durchschnittliche Anzahl induzierter Gene in den Kontrollgruppen

$s$  = Standardabweichung (vom Mittelwert der Kontrollgruppen)

$n$  = Anzahl Kontrollgruppen

### 2.18 Test auf Normalverteilung mithilfe des Kolmogorov-Smirnov-Testes

Um sicherzustellen, dass die Anzahl ABA-induzierter Gene in den Kontrollgruppen nicht signifikant von einer Normalverteilung abweicht, wurde der Kolmogorov-Smirnov-Test durchgeführt, wie in „Biostatistical Analysis“ [Zar, 1984] in Kapitel 5.10 (S. 53) beschrieben. Der Kolmogorov-Smirnov-Test wurde gewählt, weil er bei der relativ kleinen Anzahl von Kontrollgruppen ( $n=5$ ) dem Chi-quadrat-Test überlegen ist.

## 3. Ergebnisse

### 3.1 Die AthaMap Datenbank ([www.Athamap.de](http://www.Athamap.de))

#### 3.1.1 Dateninhalt von AthaMap

Die AthaMap Datenbank wurde als relationale Datenbank auf einem MSSQL-Serversystem erstellt (zur Datenbankstruktur, siehe Material- und Methodenteil, Abschnitt 2.7). Die AthaMap Datenbank enthält insgesamt mehr als  $7.4 \times 10^6$  potenzielle TFBS von 36 TFs aus 16 Familien [Steffens et al., 2005]. Diese wurden mithilfe von 42 Alignmentmatrizen detektiert und in AthaMap annotiert. In Tabelle 3 (siehe nächste Seite) ist die Anzahl potenzieller TFBS, die mit diesen Matrizen detektiert wurden (Abschnitt 2.2), aufgelistet.

Außerdem wurden in der AthaMap Datenbank die Positionen von mehr als 180.000 potenziellen kombinatorischen Elementen gespeichert, die auf Basis bekannter TF-Interaktion oder Synergismen annotiert wurden (Tab. 4, Seite 45).

Für die Detektion von potenziellen kombinatorischen Elementen nach Benutzervorgaben wurde ein Webtool entwickelt, das die Suche nach colokalisierenden TFBS ermöglicht (siehe Abschnitt 3.1.4). Dieses Webtool wurde benutzt, um kombinatorische Elemente aus den TFBS interagierender TFs zu detektieren und annotieren.

### 3. Ergebnisse

Tabelle 3: Auflistung der TFs, deren potenzielle Bindungsstellen in AthaMap gespeichert vorliegen. Für jeden TF sind die Anzahl der TFBS sowie die Referenz für die verwendete Alignmentmatrize aufgeführt.

TF	Familie	Spezies	Anzahl potenzieller TFBS	Referenz für Alignmentmatrize
ABF1	bZIP	<i>Arabidopsis thaliana</i>	2 419	[Choi et al., 2000]
bZIP910[1]	bZIP	<i>Antirrhinum majus</i>	345	[Martinez-Garcia et al., 1998]
bZIP910[2]	bZIP	<i>Antirrhinum majus</i>	470	[Martinez-Garcia et al., 1998]
bZIP911[1]	bZIP	<i>Antirrhinum majus</i>	123	[Martinez-Garcia et al., 1998]
bZIP911[2]	bZIP	<i>Antirrhinum majus</i>	145	[Martinez-Garcia et al., 1998]
TGA1	bZIP	<i>Arabidopsis thaliana</i>	53 494	[Schindler et al., 1992]
TGA1a	bZIP	<i>Nicotiana tabacum</i>	142 072	[Izawa et al., 1993]
O2	bZIP	<i>Zea mays</i>	173 685	[Izawa et al., 1993]
PIF3[1]	bHLH	<i>Arabidopsis thaliana</i>	1 154	[Martinez-Garcia et al., 2000]
PIF3[2]	bHLH	<i>Arabidopsis thaliana</i>	951	[Martinez-Garcia et al., 2000]
DOF2	DOF	<i>Zea mays</i>	1 840 355	[Yanagisawa & Schmidt, 1999]
AG	MADS	<i>Arabidopsis thaliana</i>	46 240	[Huang et al., 1993]
AGL3	MADS	<i>Arabidopsis thaliana</i>	73 298	[Huang et al., 1995]
AGL15	MADS	<i>Arabidopsis thaliana</i>	262 900	[Tang & Perry, 2003]
ABI4[1]	AP2/EREBP	<i>Zea mays</i>	12 830	[Niu et al., 2002]
ABI4[2]	AP2/EREBP	<i>Zea mays</i>	11 955	[Niu et al., 2002]
ANT	AP2/EREBP	<i>Arabidopsis thaliana</i>	294	[Nole-Wilson & Krizek, 2000]
RAV1[1]	AP2/EREBP	<i>Arabidopsis thaliana</i>	310 767	[Kagaya et al., 1999]
RAV1[2]	AP2/EREBP	<i>Arabidopsis thaliana</i>	229 983	[Kagaya et al., 1999]
TEIL	AP2/EREBP	<i>Nicotiana tabacum</i>	602 300	[Kosugi & Ohashi, 2000]
AtMYB15	MYB	<i>Arabidopsis thaliana</i>	209	[Romero et al., 1998]
AtMYB77	MYB	<i>Arabidopsis thaliana</i>	17 836	[Romero et al., 1998]
AtMYB84	MYB	<i>Arabidopsis thaliana</i>	231	[Romero et al., 1998]
CDC5	MYB	<i>Arabidopsis thaliana</i>	11 574	[Hirayama & Shinozaki, 1996]
GAMYB	MYB	<i>Hordeum vulgare</i>	315 722	[Gubler et al., 1999]
MYB.PH3[1]	MYB	<i>Petunia hybrida</i>	8 529	[Solano et al., 1995]
MYB.PH3[1]	MYB	<i>Petunia hybrida</i>	7 638	[Solano et al., 1995]
P	MYB	<i>Zea mays</i>	210 035	[Grotewold et al., 1994]
GT1	Trihelix	diverse Spezies	1 439 744	[Ouwerkerk et al., 1999]
PCF2	TCP	<i>Oryza sativa</i>	37 373	[Kosugi & Ohashi 2002]
PCF5	TCP	<i>Oryza sativa</i>	14 090	[Kosugi & Ohashi 2002]
HVH21	HD-Kn	<i>Hordeum vulgare</i>	526 877	[Krusell et al., 1997]
ALFIN1	HD-PHD	<i>Medicago sativa</i>	546 159	[Bastola et al., 1998]
ATHB1	HD-ZIP	<i>Arabidopsis thaliana</i>	66 460	[Sessa et al. 1993]
ATHB5	HD-ZIP	<i>Arabidopsis thaliana</i>	7 115	[Johannesson et al., 2001]
ATHB9	HD-ZIP	<i>Arabidopsis thaliana</i>	303	[Johannesson et al., 2001 (2)]
HAHB4	HD-ZIP	<i>Helianthus annuus</i>	90 825	[Palena et al., 1999]
AGP1	GATA	<i>Nicotiana tabacum</i>	108 199	[Sugimoto et al., 2003]
ZAP1	WRKY	<i>Arabidopsis thaliana</i>	4 302	[de Pater et al., 1996]
ID1	C2H2 (Zn)	<i>Zea mays</i>	156 641	[Kozaki et al., 2004]
TBP		diverse Spezies	16 265	[Shahmuradov et al., 2003]
CBF		diverse Spezies	61 964	[Shahmuradov et al., 2003]
Gesamt Anzahl:			7 413 949	

Die TFBS der oben aufgeführten TFs wurden durch Screening mit dem von PATSER festgelegten Threshold identifiziert (Abschnitt 2.3). Ausnahmen bilden die allgemeinen TFs TBP und CBF, die mithilfe positioneller Einschränkungen annotiert wurden (Abschnitt 2.11).

### 3. Ergebnisse

Neben den TFBS der oben aufgeführten TFs wurden auch kombinatorische Elemente in AthaMap annotiert. Die Parameter für die Detektion der Elemente wurden den in Tabelle 4 aufgeführten Referenzen entnommen. Eine Sonderstellung nimmt hier der TF RAV1 ein, der zwei DNA-Bindungsdomänen besitzt, die beide an die DNA-Bindung des TFs teilnehmen. Folgende potenzielle kombinatorische Elemente wurden mithilfe der Colokalisationsfunktion detektiert und in AthaMap annotiert (Tab. 4):

Tabelle 4: Auflistung der in AthaMap enthaltenen putativen kombinatorischen Elemente. Für jedes Element sind die zur Detektion verwendeten TFBS, die Distanz sowie die Anzahl identifizierter Elemente angegeben.

<b>Kombinatorisches Element</b>	<b>Distanz zwischen TFBS</b>	<b>Verwendete TFBS</b>	<b>Anzahl annotierter Elemente</b>	<b>Referenz</b>
RAV	3 - 20 bp	RAV1[1]/RAV1[2]	28 535	Kagaya et al., 1999
bZIP/bZIP	10 - 15 bp	TGA1a/TGA1a	1 037	Krawczyk et al. 2002
bZIP/DOF	3 - 40 bp	TGA1a/DOF2	84 389	Chen & Singh, 1999
MYC/MYB	3 - 40 bp	TGA1a/alle MYB	38 065	Baudry et al. 2004
MADS/MADS	10 - 100 bp	alle MADS/alle MADS	31 133	Trobner et al., 1992
Anzahl kombinatorischer Elemente:			183 159	

### 3. Ergebnisse

#### 3.1.2 Navigation in AthaMap

Die Darstellung des Datenbankinhaltes von AthaMap erfolgt mithilfe eines Webinterfaces unter der Internetadresse: [www.AthaMap.de](http://www.AthaMap.de). Das Webinterface (Abb. 21) wurde für die wissenschaftliche Benutzergemeinde erstellt [Steffens et al., 2004]. Die Navigation durch die einzelnen Unterseiten erfolgt durch das Menü auf der linken Seite.

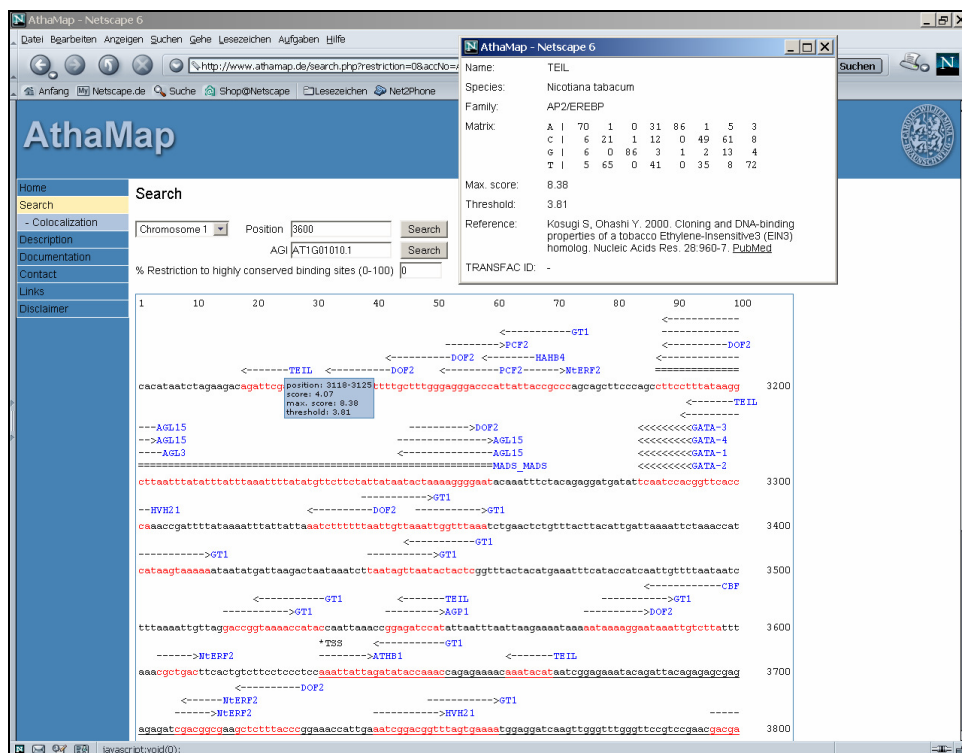


Abb. 21: Das Webinterface der AthaMap Datenbank. Über die Navigationsleiste links sind die einzelnen untergeordneten Seiten wie Beschreibung und Dokumentation von AthaMap erreichbar.

Die „Home“-Seite ist die AthaMap Startseite, auf der eine kurze Zusammenfassung des Projekts sowie Hintergrundinformationen zur AthaMap-Datenbank angezeigt werden.

Die Visualisierung von putativen TFBS ist unter der „search“-Seite zugänglich. Auf dieser Seite wählt der Benutzer entweder einen genomischen Bereich durch Eingabe einer Position und Wahl des entsprechenden Chromosoms oder gibt eine AGI-Nummer in das entsprechend bezeichnete Feld ein. Die Suche wird durch Druck auf den „Search“-Knopf gestartet.

Auf der Unterseite „Description“ erhält der Benutzer eine Beschreibung aller anderen HTML-Seiten von AthaMap, auf der „Documentation“-Unterseite findet der Benutzer

eine Auflistung aller Referenzen, aus denen die verwendeten Matrizen extrahiert wurden. Die „Contact“-Seite enthält die nötigen Informationen, um die Autoren von AthaMap zu kontaktieren. Auf der „Links“-Seite sind die wichtigsten *Arabidopsis*-Datenbanken verlinkt, namentlich TAIR, TIGR, MIPS, DATF und TRANSFAC. Auf der „Disclaimer“-Seite sind die Benutzungsregeln für die AthaMap Datenbank aufgeführt. Der Benutzer muss mit diesen Regeln einverstanden sein, um die AthaMap Datenbank nutzen zu dürfen.

#### 3.1.3 Visualisierung der *Arabidopsis*-Genomsequenz in AthaMap

Nach Eingabe einer genomischen Position oder einer AGI und Druck auf den „search“-Knopf wird der gesuchte Sequenzbereich mit den putativen TFBS im unteren Bereich des Fensters dargestellt (Abb. 21). Bei Eingabe einer AGI wird der Startpunkt des zugehörigen Gens mit den up- und downstream gelegenen 500bp Sequenz angezeigt, bei Eingabe einer Position werden 500bp up- und downstream von der eingegebenen Position dargestellt.

Der transkribierte Bereich eines Gens ist unterstrichen dargestellt, der Transkriptionsstartpunkt ist, soweit für das Gen bekannt, mit einem Stern (\*) und der Bezeichnung „TSS“ markiert.

Die Sequenz der TFBS ist in Rot dargestellt, mit einem Pfeil über der Sequenz, um die Länge und die Orientierung der TFBS anzuzeigen. Der Name des potenziell bindenden TFs erscheint neben dem Pfeil als Hyperlink. Durch einen Mausklick auf diesen Link erhält der Benutzer zusätzliche Informationen über den TF, beispielsweise die zum Screening verwendete Matrize (Abb. 21; TEIL), Threshold und max. Score, Referenz zur Matrize sowie ihre TRANSFAC-Acc., soweit vorhanden. Wird der Mauszeiger über den Pfeil der TFBS bewegt, so wird ein so genanntes Tool-Tip-Fenster angezeigt, in dem Threshold, max.Score der zugrunde liegenden Matrize sowie individuelles Score der Bindungsstelle dargestellt werden (Abb. 21). Dies ermöglicht dem Benutzer die Einschätzung der Qualität der entsprechenden TFBS. Liegt der individuelle Score näher am max. Score, so ist die Site stärker konserviert, liegt der Score näher am min. Score so ist die Sequenzkonservierung geringer.



### 3. Ergebnisse

In AthaMap annotierte potenzielle kombinatorische Elemente werden durch einen Doppelstrich über der Sequenz dargestellt. Sie besitzen weder Score noch Orientierung. Im Tool-Tip-Fenster werden lediglich Start- und Endposition angezeigt. Durch einen Mausklick auf den Namen eines kombinatorischen Elementes öffnet sich ein neues Fenster mit dem Namen des Elementes, der beim Screening tolerierten Distanz zwischen den einzelnen Bindungsstellen des Elementes, den zur Detektion der Einzelbindungsstellen verwendeten Matrizen sowie der Gesamtanzahl der detektierten kombinatorischen Elemente des entsprechenden Typs.

#### 3.1.4 Webtools in AthaMap: Restriktion von TFBS auf stark konservierte Bindungsstellen und Colokalisationsanalyse

Es ist möglich, die im Visualisierungsfenster von AthaMap dargestellten TFBS auf höher als Threshold-Score konservierte TFBS einzuschränken [Steffens et al., 2005]. Um dies zu erreichen, kann der Benutzer in dem Textfeld unterhalb der „Search“-Knöpfe eine Restriktion in Prozent [%] eingeben (Abb. 21). Die eingegebene Prozentzahl wird auf die Differenz zwischen Threshold-Score und Maximalscore einer Matrice angewendet; besitzt eine Matrice A aufgrund ihrer Nukleotidkomposition einen Threshold-Score von 5 und einen Maximalscore von 11, so beträgt die Differenz zwischen beiden Scores 6. Gibt der Benutzer in das „Restriction“-Feld eine „20“ ein, so werden TFBS der Matrice A mit einem Score von „5“ nicht mehr dargestellt. Um dargestellt zu werden, müsste eine TFBS der Matrice A einen individuellen Score von:

$$5 + (0,2 * 6) = 6,2$$

aufweisen. Diese Art der Restriktion für die Darstellung von TFBS im AthaMap Sequenzfenster ist vorteilhaft, weil sie auf alle Matrizen universell anwendbar ist, ohne den Threshold- und Maximalscore jeder Matrice zu kennen oder individuell einstellen zu müssen.

Die in AthaMap annotierten kombinatorischen Elemente besitzen keinen Score und sind von der oben beschriebenen Einschränkung nicht betroffen.

### 3. Ergebnisse

Die „Colocalization“-Unterseite der „Search“-Seite ermöglicht es dem Benutzer, nach eigenen Vorgaben Colokalisationsanalysen von TFBS durchführen zu lassen (Abb. 22). Zunächst wählt der Benutzer zwei TFs aus, deren Bindungsstellen analysiert werden sollen (Abb. 22, „TGA1a“). Nach Auswahl eines TFs wird der minimale und maximale Score für die entsprechende Matrize rechts neben dem Auswahlfenster angezeigt, und der Benutzer hat die Möglichkeit, durch Eingabe eines Scores innerhalb des angezeigten Bereiches die Anzahl der TFBS für die Analyse weiter einzuschränken. Bei Eingabe eines restriktiven Scores werden nur TFBS mit einem individuellen Score, der größer oder gleich dem benutzerdefinierten Score ist, für die Analyse verwendet.

Um die Systemstabilität zu gewährleisten, wurde für einige TFs die Anzahl der TFBS bei Colokalisationsanalysen auf 200.000 eingeschränkt.

Position		Chromosome	Orientation		Spacer
Factor 1	Factor 2		Factor 1	Factor 2	
83354	83343	1	-->	-->	11
83354	83343	1	-->	<--	11
83354	83343	1	<--	-->	11
83354	83343	1	<--	<--	11
84492	84479	1	<--	-->	13

Abb. 22: Das AthaMap Webinterface für die Durchführung von Colokalisationsanalysen

Nach Auswahl der beiden zu analysierenden Faktoren kann der Benutzer den minimalen und maximalen Abstand zwischen colokalisierenden TFBS festlegen. Im Webinterface der öffentlich zugänglichen AthaMap ([www.AthaMap.de](http://www.AthaMap.de)) ist der

maximale Abstand zwischen zwei Elementen aus Gründen der Rechenzeitausnutzung auf 50bp festgesetzt. Für eigene Analysen wurde ein Webinterface ohne diese Limitierung erstellt (Abschnitt 2.5).

Bei der Auswahl zweier Faktoren, die potenziell ähnliche Sequenzen erkennen, kann es sinnvoll sein, einen Mindestspacer einzugeben, um Überlappungen von colokalisierenden TFBS zu vermeiden. Direkt unterhalb der Textfelder für die Spacerlänge befinden sich Drop-down-Boxen (Abb. 22, „Chromosome“, „Position“), in denen insgesamt drei Sortierkriterien für die Ausgabe der Ergebnismenge ausgewählt werden können. Die Analyse wird durch einen Mausklick auf den „Search“-Knopf gestartet.

Die Ausgabe des Ergebnisses der Colokalisationsanalyse erfolgt im selben Fenster (Abb. 22, unten). Zunächst wird eine Zusammenfassung angezeigt, in der die Anzahl der TFBS von Faktor 1 und Faktor 2 in der Datenbank angegeben ist, wie viele TFBS jeweils in die Analyse miteinbezogen wurden, sowie die Anzahl der detektierten Colokalisationen. Durch einen Mausklick auf den Link „Show overview“ wird eine Tabelle eingeblendet, in der übersichtsweise die für jede Spacerlänge auftretende Anzahl an Colokalisationen angegeben ist. Wenn die Checkbox „Show zero rows in overview“ aktiviert ist, so werden auch Spacerlängen ausgegeben, für die keine Colokalisationen gefunden wurden. Dies kann bei der Weiterverwendung der Daten in Excel-Tabellen sinnvoll sein, wenn es nicht wünschenswert ist, einzelne Null-Werte auszulassen.

Unter der Übersichtstabelle befindet sich die eigentliche Ergebnistabelle mit den Positionen der colokalisierenden TFBS. Angegeben sind in jeder Tabellenzeile die Position des gewählten Faktor 1 und Faktor 2, das Chromosom, auf dem sich die Bindungsstellen befinden sowie die Länge des Spacers. Die jeweiligen Orientierungen der einzelnen Bindungsstellen einer Colokalisation sind durch Pfeile in der Tabelle dargestellt.

#### 3.2 Identifizierung putativ funktioneller TATA-Boxen

Wie in der Einleitung unter Abschnitt 1.3.1 beschrieben, spielt der TF TBP eine besondere Rolle bei der Rekrutierung des Transkriptions-Initiationskomplexes zum TSS, und legt zusammen mit dem Initiator-Element die Position des TSS fest [Zhu et al., 1995]. Diese Eigenschaften machen funktionelle TATA-Boxen zu idealen Partnern für Colokalisationsanalysen mit anderen TFs. In der aktuellen Version der AthaMap Datenbank wurden 16.265 putativ funktionelle TATA-Boxen annotiert [Steffens et al., 2005]. Die Annotation erfolgte mithilfe positioneller Kriterien, die in den folgenden Abschnitten erläutert werden.

##### 3.2.1 Das Genom von *Arabidopsis thaliana* enthält mehr TATA-Boxen als erwartet

Mit der verwendeten TBP-Matrize [Shahmuradov et al., 2003] wurden im *Arabidopsis*-Genom 247.488 potenzielle TATA-Boxen detektiert. Gemäß den aktuellen TIGR Annotationsdaten besitzt *Arabidopsis* allerdings lediglich ungefähr 30.000 Gene, die detektierte Anzahl von potenziellen TBP-Bindungsstellen ist also um ein vielfaches zu hoch.

##### 3.2.2 TATA-Boxen kumulieren in upstream-Bereichen annotierter Gene

Um festzustellen ob es „HotSpots“ gibt, an denen TATA-Boxen gehäuft auftreten, wurde die Lokalisierung jeder TFBS relativ zum nächsten Start-ATG bestimmt (Abschnitt 2.8). Das ATG wurde als Bezugspunkt gewählt, weil seine Position für alle annotierten Gene bekannt ist. Die daraus resultierende Excel-Tabelle ist in Abb. 23 graphisch dargestellt.

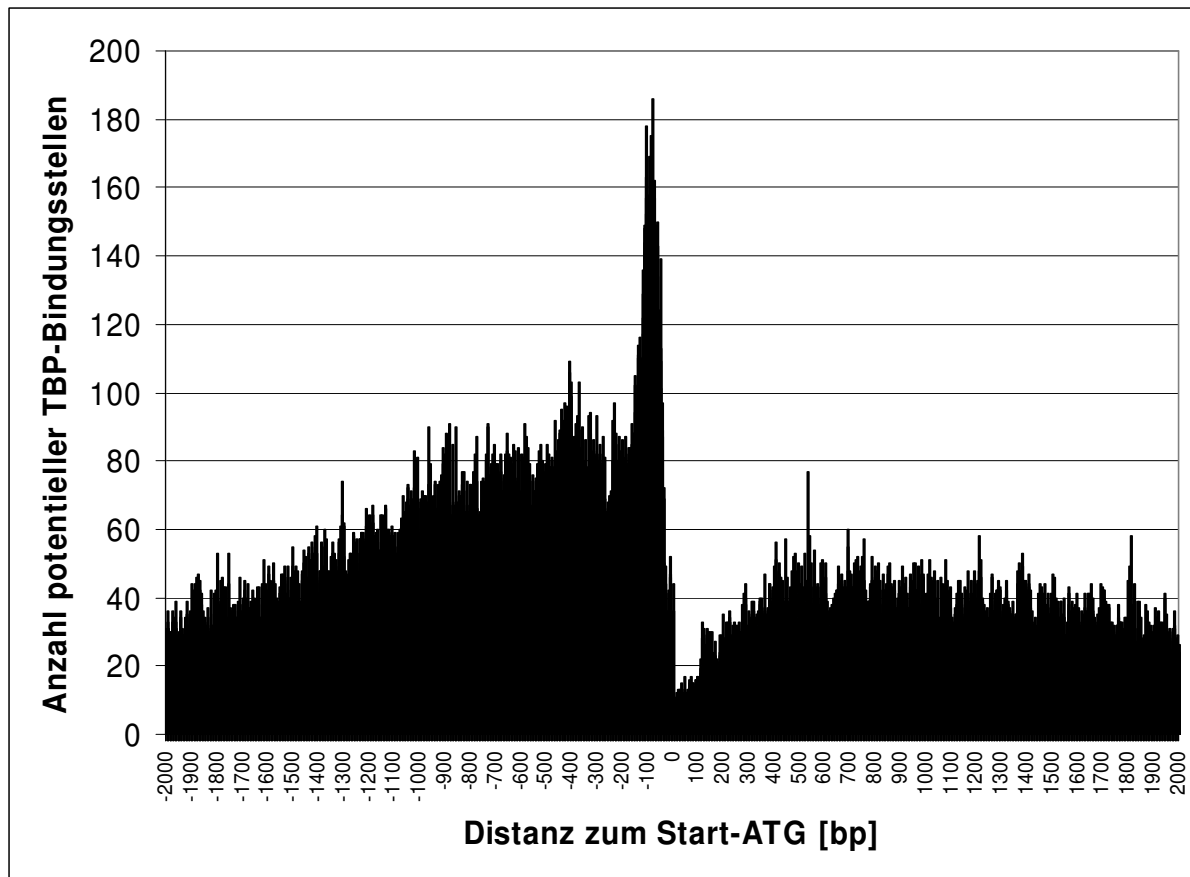


Abb. 23: Lokalisierung von potenziellen TBP-Bindungsstellen relativ zum Start-ATG (0).

Die Graphik in Abb. 23 zeigt deutlich, dass es einen stark bevorzugten Abstand von TATA-Boxen relativ zum ATG gibt. Bis zu einem Abstand von ca. 100bp im Bereich upstream des ATG treten TBP-TFBS gehäuft auf. Das Maximum der Häufung liegt bei 79bp upstream vom ATG (nicht gezeigt). Außerdem ist festzustellen, dass TBP-TFBS generell häufiger im upstream-Bereich von Genen auftreten als downstream vom Start-ATG. Bei ca. -300bp im upstream-Bereich tritt ein deutlicher Einschnitt auf (Abb. 23).

#### 3.2.3 TATA-Boxen halten einen bevorzugten Abstand von $35\text{bp} \pm 5\text{bp}$ zum TSS ein

Es ist bekannt, dass sich die TATA-Box in vielen eukaryontischen Promotoren in einem Abstand von ca. 30 bp upstream des Transkriptionsstartpunktes (TSS) befindet [Burley & Roeder 1996]. Um festzustellen, ob dies auch für die TBP-Bindungsstellen in *Arabidopsis thaliana* zutrifft, wurde als nächster Schritt die Lokalisierung von TBP-TFBS relativ zum TSS analysiert. Zu diesem Zweck wurde

### 3. Ergebnisse

eine Koloalisationsanalyse von TATA-Boxen und TSS im Abstand von 0-100bp durchgeföhrt (Abschnitt 2.5) und graphisch dargestellt (Abb. 24).

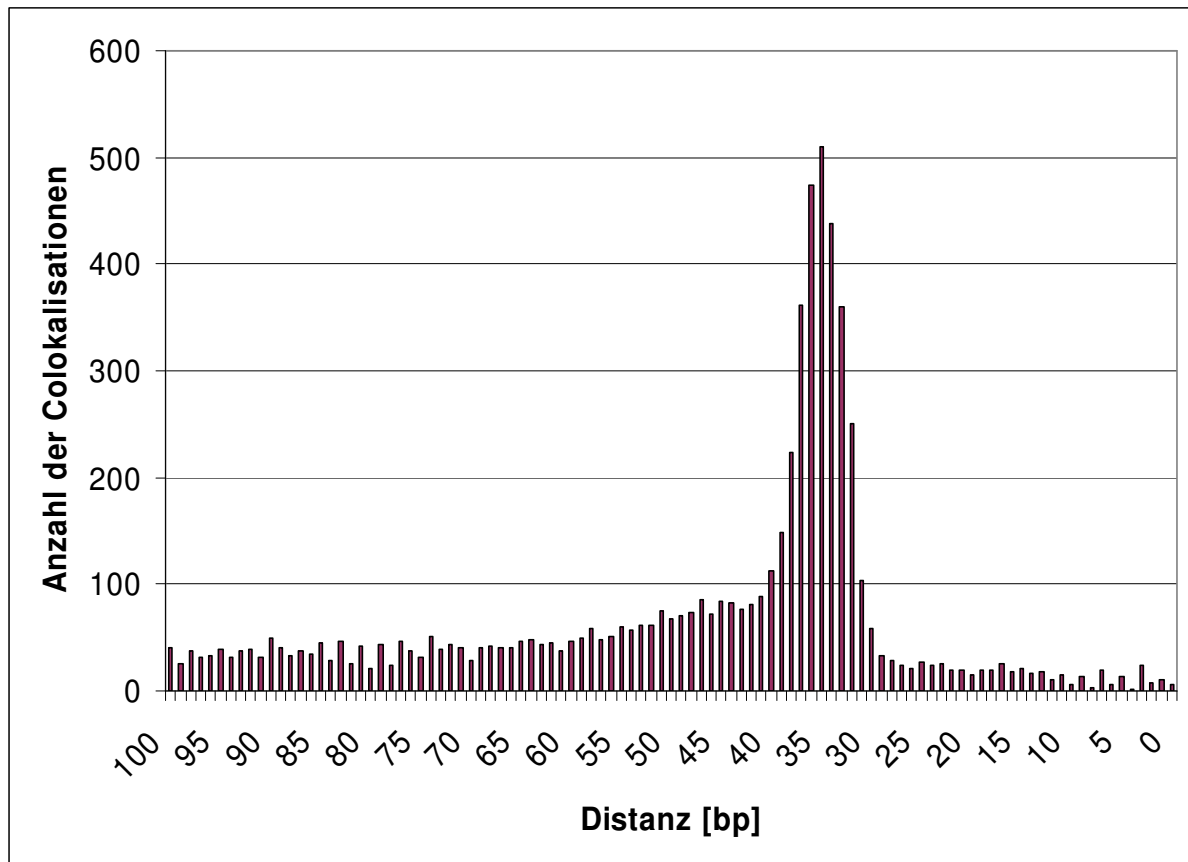


Abb. 24: Lokalisierung von TBP-TFBS relativ zum nächsten TSS. Als Startpunkt einer TBP-Bindungsstelle wird hier das vom Genstart aus gesehen distale Nukleotid betrachtet. Das Maximum des Graphen liegt bei 35bp upstream vom TSS.

Wie in Abb. 24 zu erkennen ist, lokalisieren TBP-Bindungsstellen bevorzugt in einem Abstand von 35bp upstream vom TSS. Insgesamt gibt es 3559 TATA-Boxen im upstream-Bereich von 2850 Genen, die im Abstand von 30-40bp mit einem TSS koloalisieren.

Um diese Eigenschaft für die positionelle Einschränkung der TATA-Boxen zu nutzen, wurden zunächst alle potenziellen TBP-Bindungsstellen in die AthaMap-Datenbank importiert, die sich im Bereich von 0bp bis 100bp upstream eines TSS befinden. Der Bereich, in dem TATA-Boxen upstream eines TSS als putativ funktionell eingestuft und annotiert werden, wurde auf 100bp angesetzt, weil die Annotation der TSS fast ausschließlich über cDNAs erfolgte [Haas et al., 2003], die nicht vollständig sein müssen. Da nicht davon ausgegangen werden kann, dass der TSS in allen Fällen punktgenau bestimmt werden konnte, wurden nicht nur TATA-Boxen mit dem

Abstand 35bp zum nächsten TSS als putativ funktionell betrachtet, sondern bis zu einem Abstand von 100bp zum nächstgelegenen TSS.

Dieses Kriterium ist allerdings nur anwendbar, wenn der TSS des entsprechenden Gens bekannt ist.

#### 3.2.4 Kriterien für die Identifizierung und Annotation von TATA-Boxen im upstream-Bereich von Genen ohne bekannten TSS

Nur von ca. der Hälfte der in *Arabidopsis* annotierten Gene (16.319 Gene) ist ein TSS bekannt, von den anderen Genen (13.674 Gene) ist nur das putative Start-ATG bekannt und annotiert. Um zu untersuchen, ob diese Gene TATA-Boxen enthalten, mussten Kriterien definiert werden. Vor dem Start-ATG befindet sich normalerweise ein 5' untranslatierter Bereich (5'UTR) variabler Länge. Im Falle der Gene mit annotierten TSS ist die Länge des 5'UTR bekannt, im Falle der Gene ohne annotierten TSS unbekannt. Die Länge des 5'UTR wurde für alle Gene mit bekanntem TSS bestimmt (Abschnitt 2.13) (Abb. 25, nächste Seite) und daraus die durchschnittliche 5'UTR-Länge errechnet. Hierbei wurde für jedes Gen das hauptsächlich auftretende Transkript, das im Regelfall die Bezeichnung „AG1.1“ trägt (Primärtranskript), zugrunde gelegt. Die durchschnittliche Länge der 5'UTR bei den in *Arabidopsis* annotierten Primärtranskripten beträgt 193bp. Mehr als 70% der Gene mit annotiertem TSS besitzen einen 5'UTR der Länge 193bp oder kürzer (Abb. 25, senkrechter Balken). Da sich die Gene mit und ohne annotierten TSS sehr wahrscheinlich nicht funktionell unterscheiden, können wir im Falle der Gene ohne annotierten TSS analog davon ausgehen, dass 70% der Gene einen 5'UTR der Länge 193bp oder kürzer besitzen.

Da der Abstand TSS-ATG im Durchschnitt 193bp lang ist und der Abstand TBP-TSS bevorzugt bei 35bp liegt, ergibt sich daraus, dass TBP-TFBS einen konservierten Abstand zum Start-ATG einhalten müssten, der sich aus diesen beiden Abständen ergibt (Abb. 26, nächste Seite).

Während die Abstände zwischen TSS und ATG sowie TBP und TSS nur bei Genen mit bekanntem TSS bestimmt werden können, kann der Abstand TBP-ATG als Kriterium für die Gene ohne annotierten TSS herangezogen werden.

### 3. Ergebnisse

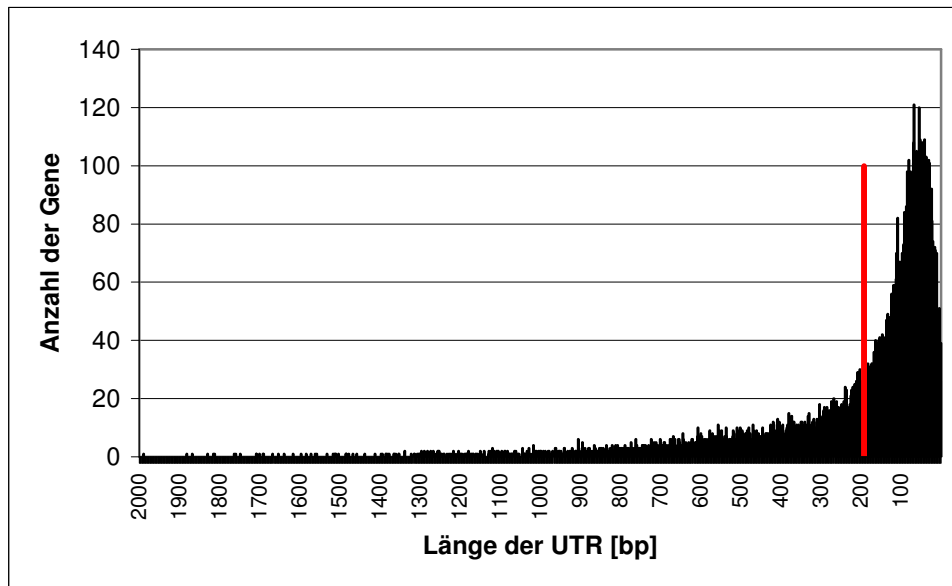


Abb. 25: Anzahl der Gene mit einer bestimmten 5'UTR-Länge. Die durchschnittliche Länge der 5'UTR (193bp) ist mit einem senkrechten Balken markiert.

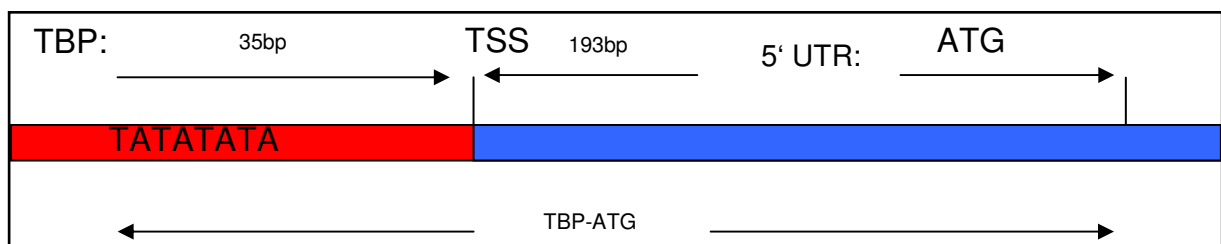


Abb. 26: Die Länge der 5'UTR und die Distanz zwischen TBP und TSS addieren sich zu dem bevorzugte Abstand der TBP-TFBS zum ATG.

Wenn für die Gene mit unbekannten TSS dieselbe UTR-Länge veranschlagt wird sowie gleichermaßen ein Toleranzfenster von 100bp für die Annotation putativ funktioneller TATA-Boxen, führt dies zu dem Annotationskriterium für die TBP-TFBS, die im upstream-Bereich von Genen ohne bekannten TSS lokalisieren. Für diese Gene wurden alle TBP-TFBS annotiert, die in einen Abstand kleiner oder gleich  $193\text{bp} + 100\text{bp} = 293\text{bp}$  upstream vom Start-ATG lokalisieren.



### 3. Ergebnisse

#### 3.2.5 Die Mehrzahl der Gene in *Arabidopsis* besitzt keine TATA-Box

Mit den oben genannten Kriterien wurden insgesamt 16.265 TATA-Boxen im *Arabidopsis*-Genom identifiziert [Steffens et al., 2005]. Die Anzahl TATA-Boxen in der Gengruppe mit TSS und ohne bekannten TSS wurde verglichen (Tabelle 5).

Tabelle 5: Die Häufigkeit des Auftretens von TATA-Boxen vor Genen mit bekanntem und unbekanntem TSS.

	Gene mit bekanntem TSS	Gene mit unbekanntem TSS	Insgesamt
<b>Anzahl der Gene</b>	16319 (100%)	13674 (100%)	29993 (100%)
<b>Gene mit mindestens einer TATA-Box</b>	4611 (28,3%)	5611 (41%)	10222 (34,1%)
<b>Gene ohne TATA-Boxen</b>	11708 (71,7%)	8063 (59%)	19771 (65,9%)

Die Analyse in Tabelle 5 zeigt, dass von 29993 Genen im *Arabidopsis*-Genom durchschnittlich bei 10222 Genen ( $\approx 34\%$ ) mindestens eine TATA-Box im upstream-Bereich des Gens detektiert werden kann. Dieser Befund steht im Einklang mit den kürzlich von Molina et al. publizierten Daten, die den Anteil TATA-Box-haltiger Promotoren im *Arabidopsis*-Genom aufgrund ihrer Analysen mit 28,8% angeben [Molina & Grotewold 2005].

Die Analyse zeigt außerdem, dass die Anteile der Gene mit TATA-Box deutlich unterschiedlich groß sind. Während in der Gengruppe ohne bekannten TSS 41% der Gene mindestens eine TATA-Box besitzen, beträgt der Anteil TATA-Box-haltiger Gene in der Gruppe mit bekanntem TSS nur lediglich 28,3%.

#### 3.2.6 Bei Genen mit unbekanntem TSS treten häufiger mehrfache TATA-Boxen auf

Um zu untersuchen, inwieweit die unterschiedlichen Anteile TATA-Box-haltiger Gene durch die unterschiedlichen Annotationskriterien bei Genen mit und ohne bekannten TSS bedingt sind, wurde die Lokalisierung von TATA-Boxen relativ zum Start-ATG in den Genen beider Gruppen untersucht (Abschnitt 2.8). Die Ergebnisse sind auf Abbildung 27 (nächste Seite) graphisch dargestellt.

### 3. Ergebnisse

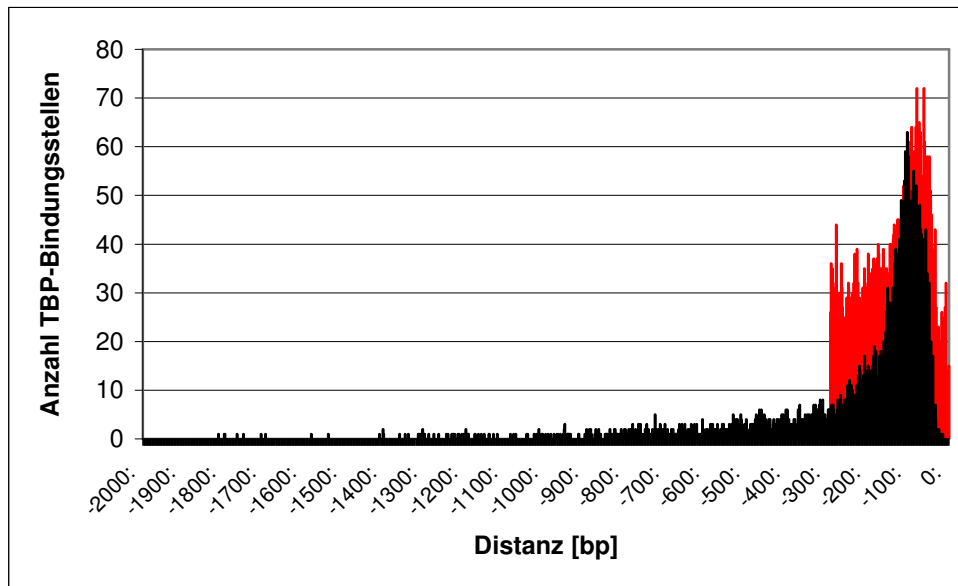


Abb. 27: Abstand von TSS-annotierten und ATG-annotierten TBP-TFBS relativ zum nächsten Start-ATG. **Grau**: ATG-annotierte TATA-Boxen. **Schwarz**: TSS-annotierte TATA-Boxen. Das ATG (0) wurde als Bezugspunkt gewählt, weil es für alle annotierten Gene bekannt ist.

Abbildung 27 zeigt eine Überlagerungsgraphik der Lokalisierung von TSS- und ATG-annotierten TATA-Boxen. Hier wird im Vergleich deutlich, dass die beiden Populationen von TBP-Matches eine sehr ähnliche Lokalisierung zum nächsten Start-ATG zeigen. In beiden Gruppen von TBP-TFBS zeigt sich eine bevorzugte Lokalisierung in einem Abstand von  $\approx 100\text{bp}$  zum nächsten ATG.

Bei den TBP-TFBS vor Genen mit bekanntem TSS liegt dieser bevorzugte Abstand genau bei 104bp upstream vom ATG, bei den Genen ohne bekannten TSS liegt das TBP-Maximum bei 79bp upstream vom ATG des nächsten Gens. Im Bereich bis -200bp vom ATG ist in beiden TBP-Populationen ein Abfall der TBP-Häufigkeit zu beobachten, der bei den Genen mit bekanntem TSS deutlich stärker ist als bei den Genen ohne bekannten TSS. Insgesamt ist die Streubreite des Abstandes bei den Genen ohne bekannten TSS größer als bei den Genen mit bekanntem TSS (Abb. 27). Gleichermäßen gibt es mehr TATA-Boxen vor den Genen ohne bekannten TSS, und damit wahrscheinlich auch mehr falsch positive TATA-Boxen.

Tabelle 6 zeigt die Ergebnisse der Analyse, in der das Auftreten mehrfacher TATA-Boxen vor Genen mit bekanntem TSS und ohne bekannten TSS verglichen wurde. Bei den Genen ohne bekannten TSS ist die Anzahl der Gene mit mehrfachen TATA-Boxen fast doppelt so groß (Tabelle 6, 2063 Gene) wie bei Genen mit bekanntem

### 3. Ergebnisse

TSS (1080 Gene). Dies könnte ein Indiz dafür sein, dass bei den Genen ohne bekannten TSS mehr falsch positive TATA-Boxen identifiziert wurden.

Tabelle 6: Häufigkeit des Auftretens multipler TATA-Boxen im upstream-Bereich von Genen mit und ohne bekannten TSS.

Anzahl der TATA-Boxen	Gene mit bekanntem TSS	Gene mit unbekanntem TSS
0	11708	8063
1	3531	3548
2	761	1282
3	187	428
4	50	152
5	30	76
6	18	21
7	7	16
8	3	15
9	4	15
10	5	9
> 10	15	49
>= 2	1080	2063

#### 3.2.7 CCAAT-Boxen besitzen keinen stark konservierten Abstand zum ATG oder TSS

Mit der CCAAT-Box Matrize und dem von PATSER gesetzten automatischen Threshold wurden 1.337.117 potenzielle CCAAT-Boxen detektiert (Abschnitt 2.2). Die CCAAT-Box ist, soweit in einem Promotor vorhanden, in ihrer Position relativ zum TSS weniger stark konserviert als die TATA-Box. Ihre Position liegt nach Untersuchungen in eukaryontischen Promotoren 60bp bis 100bp upstream vom TSS [Mantovani, 1999]. Im Rahmen dieser Arbeit wurde untersucht, inwieweit die TFBS, die mit der CCAAT-Matrize [Shahmuradov et al., 2003] detektiert wurden, einen bestimmten Abstand zum TSS bevorzugen. Die Lokalisierung der CCAAT-Boxen relativ zum ATG wurde mit den TFBSitesTools bestimmt, wie unter Material und Methoden beschrieben (Abschnitt 2.8). Die Lokalisierung der CCAAT-Boxen relativ zum TSS wurde mithilfe von Colokalisationsanalysen bestimmt (Abschnitt 2.5). Die Lokalisierung von CCAAT-Boxen relativ zum TSS ist in Abb. 28 graphisch dargestellt.

### 3. Ergebnisse

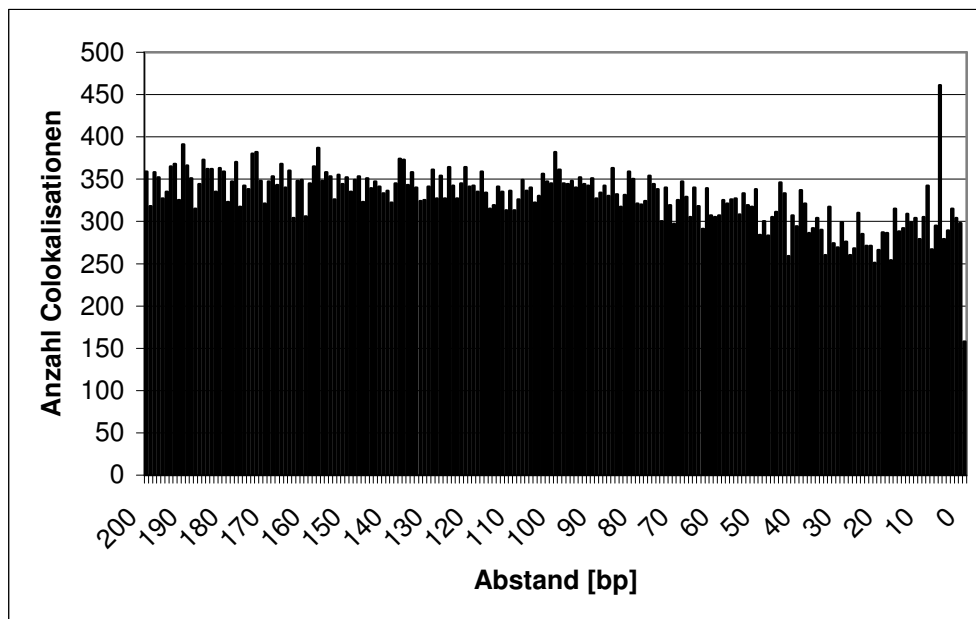


Abb. 28: Lokalisierung der CCAAT-Box relativ zum nächsten TSS.

Es ist keine Bevorzugung eines gehäuft auftretenden Abstandes oder Abstandsbereiches relativ zum TSS erkennbar. Auch eine Positionsanalyse aller anfänglich detektierten CCAAT-Boxen in einem wesentlich größeren Abstandsbereich (-2000bp bis +2000bp) relativ zum nächsten Start-ATG zeigt ebenfalls keine bevorzugte Lokalisierung (Abb. 29, nächste Seite). In der Graphik treten zwei einzelne Peaks auf, und zwar bei -7bp und +5bp relativ zum Nullpunkt (ATG). Diese Peaks sind auf die Überlagerung von Start-ATG und der CCAAT-Box Matrize zurückzuführen (Abb. 29). Die CCAAT-Box Matrize fordert an Position 8 und Position 9 die Sequenz „AT“, und da diese Nukleotide Bestandteil des ATG-Startcodons sind, gibt es in Position -7 relativ zum ATG besonders viele potenzielle CCAAT-Box-Bindungsstellen. Dasselbe gilt natürlich für CCAAT-Box-Matches auf dem reversen DNA-Strang.

Da jedoch die Funktionalität von CCAAT-Boxen wie auch bei den TATA-Boxen im Bereich von -60 bis -100bp relativ zum TSS nachgewiesen wurde [Mantovani, 1999], wurden dieselben Kriterien für die positionelle Einschränkung der CCAAT-Boxen angelegt wie für die Einschränkung der TATA-Boxen. CCAAT-Boxen wurden im Bereich 100bp upstream eines TSS annotiert bzw. im Bereich 293bp upstream von Genen ohne bekanntem TSS.

### 3. Ergebnisse

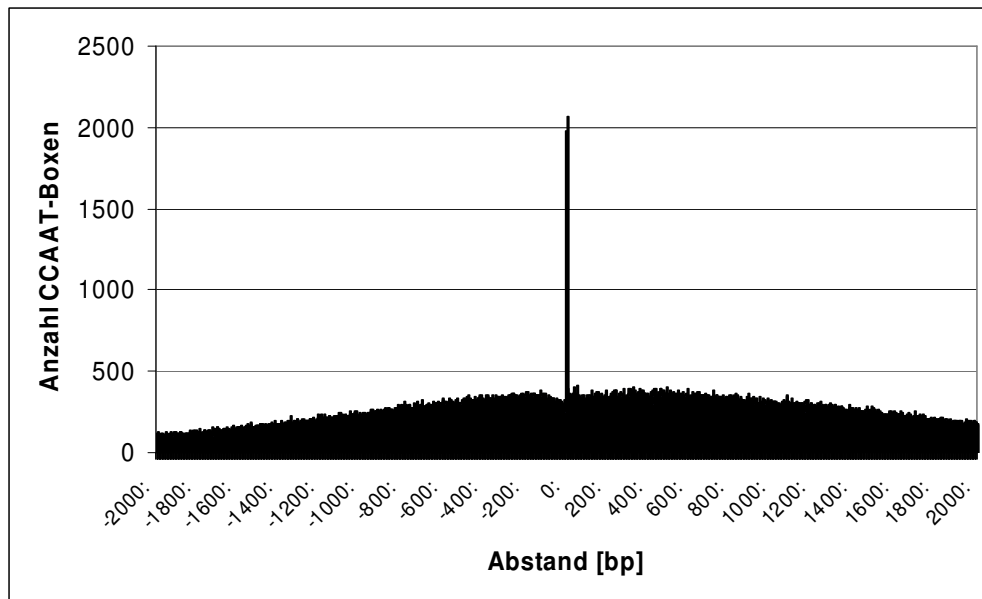


Abb. 29: Lokalisierung der CCAAT-Box relativ zum nächsten Start-ATG

### 3.3 Prädiktion von potenziell ABA-regulierten Genen anhand von Colokalisationsanalysen

#### 3.3.1 Ein Großteil der Gene mit intergenischen ABF1-TBP-Colokalisationen ist ABA-induzierbar

ABF1 ist ein potenzieller Regulator von ABA-induzierten Stressantworten und ein möglicher Regulator von ABREs [Choi et al., 2000; Kim et al., 2004]. Um eine potenzielle Interaktion zwischen dem bZIP-TF ABF1 und dem TBP-Protein zu untersuchen, wurde eine Colokalisationsanalyse (Abschnitt 2.5) mit ABF1- und TBP-Bindungsstellen in einem Abstand von 0bp bis 50bp durchgeführt. Mithilfe des Localizer-Programms (Abschnitt 2.10) wurden diejenigen Gene identifiziert, bei denen die ABF1-TBP-Colokalisation intergenisch lokalisiert ist. Insgesamt wurden 42 Genen identifiziert, bei denen ABF1 und TBP upstream des transkribierten Bereiches liegen. Die AGI-Nummern der 42 Gene sowie die putative Funktion sind in Tabelle 7 aufgeführt. Von den 42 Genen sind 35 Gene auf dem ATH1-Chip (Affymetrix) repräsentiert. Die Überprüfung des Expressionsstatus (Abschnitt 2.15) zeigte, dass bei Stimulation der Pflanzen mit ABA 21 der 35 Gene induziert vorliegen, zwei Gene reprimiert vorliegen und die Transkription der restlichen 12 Gene nicht verändert ist (Tabelle 7).

Tabelle 7: Gene der Prädiktionsgruppe mit ABF1-TBP-Colokalisation im Abstand 0-50bp  
Expressionsstatus: n.d. : nicht detektiert (keine Änderung); >1 = induzierte Gene; <1 = reprimierte Gene

AGI	Annotation	Expressionsstatus bei ABA-Stimulation
<b>At1g03030</b>	peroxidase, putative, similar to peroxidase; peroxidase ATP14a (Arabidopsis thaliana) gi 1546690 emb CAA67335	1,266
<b>At1g04820</b>	tubulin alpha-2/alpha-4 chain (TUA4), nearly identical to SP:P29510 Tubulin alpha-2/alpha-4 chain from (Arabidopsis thaliana)	n.d.
<b>At1g07590*</b>	pentatricopeptide (PPR) repeat-containing protein, low similarity to DNA-binding protein (Triticum aestivum) GI:6958202; contains Pfam profile PF01535: PPR repeat	n.d.
<b>At1g10590**</b>	DNA-binding protein-related, contains weak similarity to G-quartet DNA binding protein 3 (Tetrahymena thermophila) gi 4583503 gb AAD25098	n.d.
<b>At1g11870</b>	seryl-tRNA synthetase, putative / serine--tRNA ligase, putative, similar to PIR T03949 serine--tRNA ligase (EC 6.1.1.11) serS {Zea mays}; contains Pfam profiles PF00587:	n.d.

### 3. Ergebnisse

	tRNA synthetase class II core domain (G, H, P, S and T), PF02403: Seryl-tRNA synthetase N-terminal domain	
<b>At1g19860</b>	zinc finger (CCCH-type) family protein, contains Pfam domain, PF00642: Zinc finger C-x8-C-x5-C-x3-H type (and similar)	1,446
<b>At1g22510</b>	zinc finger (C3HC4-type RING finger) family protein, contains Pfam profile: PF00097 zinc finger, C3HC4 type (RING finger)	1,157
<b>At1g22600</b>	hypothetical protein	0,832
<b>At1g26670</b>	vesical transport v-SNARE 12 (VTI12) / vesicle soluble NSF attachment protein receptor VTI1b (VTI1B) receptor VTI1b, identical to SP Q9SEL5 Vesicle transport v-SNARE 12 (AtVTI12) (Vesicle transport v-SNARE protein VTI1b) (Vesicle soluble NSF attachment protein receptor VTI1b) (AtVTI1b) {Arabidopsis thaliana}	1,594
<b>At1g48830</b>	40S ribosomal protein S7 (RPS7A), similar to 40S ribosomal protein S7 homolog GI:5532505 from (Brassica oleracea)	n.d.
<b>At1g51400</b>	photosystem II 5 kD protein, 100% identical to GI:4836947 (F5D21.10)	n.d.
<b>At1g53165</b>	protein kinase, putative similar to serine/threonine protein kinase 24 [Homo sapiens]	n.d.
<b>At1g53170</b>	encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-8). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole.	3,886
<b>At1g58340</b>	MATE efflux protein-related, contains Pfam profile: PF01554 uncharacterized membrane protein family UPF0013	1,655
<b>At1g76570</b>	chlorophyll A-B binding family protein, similar to chlorophyll A-B binding protein GB:P12470 (Nicotiana plumbaginifolia); contains Pfam profile: PF00504 Chlorophyll A-B binding proteins	1,661
<b>At2g04100</b>	MATE efflux family protein, similar to ripening regulated protein DDTFR18 (Lycopersicon esculentum) GI:12231296; contains Pfam profile PF01554: Uncharacterized membrane protein family	1,225
<b>At2g14740</b>	vacuolar sorting receptor, putative, nearly identical to vacuolar sorting receptor homolog (Arabidopsis thaliana) GI:1737220; contains a calcium-binding EGF-like domain signature	1,233
<b>At2g22240</b>	inositol-3-phosphate synthase isozyme 2 / myo-inositol-1-phosphate synthase 2	6,767
<b>At3g03250</b>	UTP--glucose-1-phosphate uridylyltransferase, putative / UDP-glucose pyrophosphorylase, putative / UGPase, putative, strong similarity to SP P19595 UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) (UDP-glucose pyrophosphorylase) (UDPGP) (UGPase) {Solanum tuberosum}; contains Pfam profile PF01704: UTP--glucose-1-phosphate uridylyltransferase	n.d.
<b>At3g04620</b>	expressed protein	n.d.
<b>At3g11330</b>	leucine-rich repeat family protein	1,467
<b>At3g18360</b>	VQ motif-containing protein contains PF05678: VQ motif	n.d.
<b>At3g23920</b>	beta-amylase, putative / 1,4-alpha-D-glucan maltohydrolase, putative, similar to beta-amylase enzyme (Arabidopsis thaliana) GI:6065749, beta-amylase PCT-BMYI from (Solanum tuberosum); contains Pfam profile PF01373: Glycosyl hydrolase family 14	6,444
<b>At3g24440</b>	fibronectin type III domain-containing protein, contains Pfam profile PF00041: Fibronectin type III domain	n.d.
<b>At3g26910</b>	hydroxyproline-rich glycoprotein family protein, common family member:At2g33490 (Arabidopsis thaliana)	1,635
<b>At3g48510</b>	expressed protein	10,557

### 3. Ergebnisse

<b>At3g48810***</b>	pentatricopeptide (PPR) repeat-containing protein, contains Pfam profile PF01535: PPR repeat	0,768
<b>At3g57520</b>	alkaline alpha galactosidase, putative, similar to alkaline alpha galactosidase II (Cucumis melo) GI:29838631; contains Pfam profile PF05691: Raffinose synthase or seed imbibition protein Sip1	3,085
<b>At3g63060</b>	circadian clock coupling factor, putative, similar to gb:AAK56924 circadian clock coupling factor ZGT {Nicotiana tabacum}	11,207
<b>At4g01940</b>	nitrogen fixation NifU-like family protein, similar to apricot NifU homolog partial CDS, GenBank accession number U95179; contains Pfam profile: PF01106 NifU-like domain	n.d.
<b>At4g16380</b>	expressed protein	1,172
<b>At5g14500</b>	aldose 1-epimerase family protein, similar to apospory-associated protein C, Chlamydomonas reinhardtii, EMBL:AF195243 Pfam profile PF01263: Aldose 1-epimerase	1,995
<b>At5g18670</b>	beta-amylase, putative (BMY3) / 1,4-alpha-D-glucan maltohydrolase, putative, almost identical to beta-amylase BMY3 GI:15149457 from (Arabidopsis thaliana); identical to cDNA putative beta-amylase BMY3 (BMY3) GI:15149456	1,715
<b>At5g25240</b>	expressed protein	1,994
<b>At5g60790</b>	ABC transporter family protein, similar to ABC transporter homolog PnATH GI:7573600 from (Populus nigra)	1,399
<b>At1g35980</b>	pseudogene, hypothetical protein, probably a fragment of the entire gene	Nicht auf dem Chip
<b>At1g67785</b>	expressed protein	Nicht auf dem Chip
<b>At3g12970</b>	expressed protein	Nicht auf dem Chip
<b>At3g18950</b>	transducin family protein / WD-40 repeat family protein, contains 7 WD-40 repeats (PF00400); similar to En/Spm-like transposon protein GI:2739374 from (Arabidopsis thaliana); no characterized homologs	Nicht auf dem Chip
<b>At3g46385</b>	expressed protein	Nicht auf dem Chip
<b>At4g12680</b>	expressed protein ; expression supported by MPSS	Nicht auf dem Chip
<b>At5g15970</b>	stress-responsive protein (KIN2) / stress-induced protein (KIN2) / cold-responsive protein (COR6.6) / cold-regulated protein (COR6.6), identical to SP P31169 Stress-induced KIN2 protein (Cold-induced COR6.6 protein) {Arabidopsis thaliana}	Nicht auf dem Chip

\* Die Gene At1g07590 und At1g07610 sind benachbart. Die Colokalisation wird aufgrund des Abstandes dem Gen At1g07590 zugeordnet.

\*\*Die Gene At1g10590 und At1g10600 sind benachbart. Die Colokalisation wird aufgrund des Abstandes dem Gen At1g10590 zugeordnet. Das Gen At1g10600 ist nicht auf dem ATH1-Chip repräsentiert.

\*\*\*At3g48800 und At3g48810 sind benachbart. Die Colokalisation wird dem Gen At3g48810 zugeordnet, weil At3g48800 nicht auf dem ATH1-Chip repräsentiert ist.

Die Untersuchung des Expressionsstatus der in Tabelle 7 aufgeführten Gene mit Genevestigator (Abschnitt 2.15) zeigt, dass der Anteil ABA-induzierter Gene relativ hoch ist (60%). Um zu untersuchen, ob in den 35 identifizierten Genen (Prädiktionsgruppe) signifikant viele ABA-induzierte Gene auftreten, werden fünf Kontrollgruppen von zufällig ausgewählten Genen (Abschnitt 2.14) auf ihre ABA-Induzierbarkeit untersucht und statistisch mit dem t-Test ausgewertet. Jede Gruppe



enthält dieselbe Anzahl Gene (35 Gene) wie die Prädiktionsgruppe mit ABF1-TBP-Colokalisationen. Die AGI-Nummern der einzelnen Gene in den Kontrollgruppen sind im Anhang (Abschnitt 7.2) aufgeführt. Die nachfolgende Abbildung 30 zeigt den graphischen Vergleich der Anzahl ABA-induzierter Gene in der Prädiktionsgruppe mit der durchschnittlichen Anzahl ABA-induzierter Gene in den Kontrollgruppen.

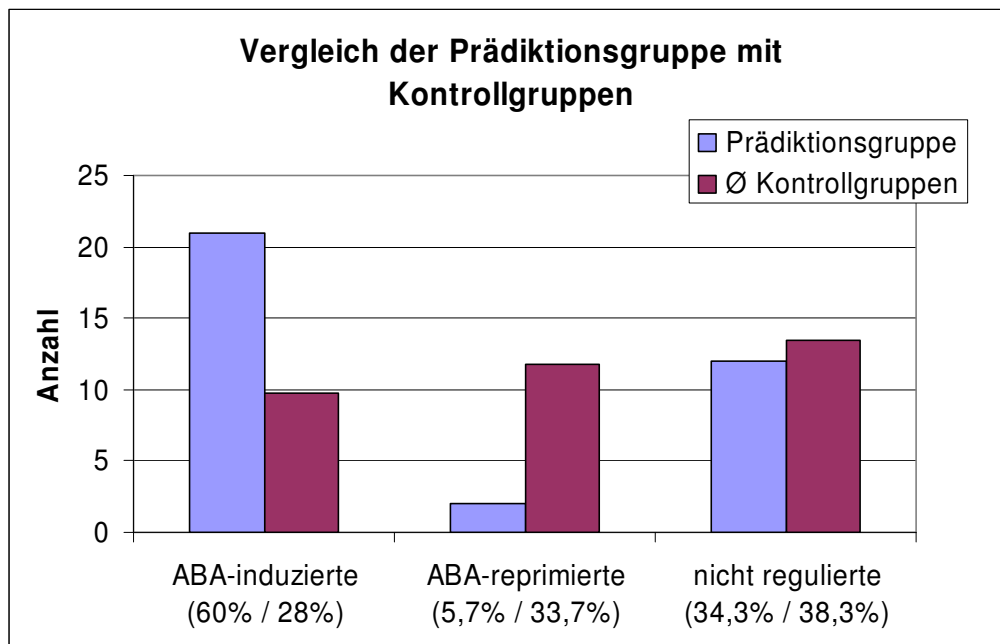


Abb. 30: Vergleich der Anzahl an ABA-induzierten, ABA-reprimierten und nicht ABA-regulierten Genen in der Prädiktionsgruppe mit den entsprechenden Durchschnittswerten für die Kontrollgruppen. In Klammern die jeweiligen prozentualen Anteile in der Prädiktionsgruppe und in den Kontrollgruppen.

Im graphischen Vergleich von Prädiktions- und Kontrollgruppen (Abb. 30) ist erkennbar, dass ABA-induzierte Gene in der Prädiktionsgruppe mehr als doppelt so häufig auftreten wie in den Kontrollgruppen. Außerdem ist der Anteil ABA-reprimierbarer Gene in der Prädiktionsgruppe erheblich geringer als in den Kontrollgruppen. In den Kontrollgruppen besitzen die durchschnittlichen Anteile ABA-induzierbarer, ABA-reprimierbarer und nicht ABA-regulierter Gene vergleichbare Größenordnungen. In der Prädiktionsgruppe hingegen stellt der Anteil ABA-induzierbarer Gene den Hauptteil der Gene dar.

Der statistische Vergleich von Prädiktionsgruppe mit den fünf Vergleichsgruppen sollte zeigen, ob in der Prädiktionsgruppe signifikant viele ABA-induzierte Gene auftreten oder ob die Anreicherung durch Zufall bedingt sein kann. Dies wurde mithilfe des t-Testes (Abschnitt 2.17) überprüft.

### 3. Ergebnisse

Beim t-Test werden zwei Hypothesen gegeneinander geprüft, zum einen die Nullhypothese ( $H_0$ ), und die entsprechende (alternative) Gegenhypothese ( $H_A$ ).

$H_0$ : Die Anzahl ABA-induzierter Gene in der Prädiktionsgruppe weicht nicht signifikant von der Anzahl induzierter Gene in den Vergleichsgruppen ab.

$H_A$ : Die Anzahl ABA-induzierter Gene in der Prädiktionsgruppe ist signifikant höher als in den Vergleichsgruppen.

Parameter für t-Test:

$N_i = 21$  (induzierte Gene in der Prädiktionsgruppe)

$R_i = 9,8$  (Ø ABA-induzierte Gene in den Kontrollgruppen)

$n = 5$  (Anzahl der Gruppen aus zufällig gewählten Genen)

$s = 2,49$  (Standardabweichung)

$$t = \frac{(21 - 9,8)}{\frac{2,49}{\sqrt{5}}} = \frac{11,2}{1,11} = 10,06$$

Abb. 31: t-Test zur Signifikanz der Häufigkeit ABA-induzierter Gene in der Prädiktionsgruppe und den Kontrollgruppen

Für einen signifikanten Unterschied zwischen Prädiktions- und Kontrollgruppen bei einem Signifikanzniveau von 0,005 muss  $t$  den kritischen Wert von 5,598 überschreiten [Zar, 1984]. Der Wert von  $t=10,06$  zeigt, dass auf einem Signifikanzniveau von 0,005 (also einer Fehlerwahrscheinlichkeit von 0,5%) die Anzahl der ABA-induzierten Gene in der Prädiktionsgruppe signifikant höher ist als in den Kontrollgruppen.

Damit die Kontrollgruppen zum Vergleich herangezogen werden können, muss sichergestellt sein, dass die Anzahl induzierter Gene in diesen Gruppen nicht signifikant von einer angenommenen Gleichverteilung abweicht.

Um festzustellen, ob die Kontrollgruppen überhaupt geeignet, also gleich verteilt sind, wird der Kolmogorov-Smirnov-Test [Zar, 1984] durchgeführt.

Der Kolmogorov-Smirnov-Test (Abschnitt 2.18) zeigt, dass die Anzahl ABA-induzierbarer Gene in den Kontrollgruppen keine signifikante Abweichung von einer Gleichverteilung zeigt und die Gruppen daher als Kontrolle geeignet sind (Daten nicht gezeigt).

#### 3.3.2 Die ABF1-Bindungsstelle von ABF1-TBP-Colokalisationen tritt bevorzugt im upstream-Bereich auf

Als nächstes sollte die ABA-Induzierbarkeit von Genen mit einer potenziellen ABF1-Bindungsstelle im transkribierten Bereich eines Gens untersucht werden. Dazu wurde eine Colokalisationsanalyse von ABF1 und TBP in einem Abstand von 100bp bis 200bp durchgeführt. Durch Analyse mit dem Localizer-Tool wurden insgesamt 16 Gene identifiziert, bei denen eine TATA-Box upstream vorhanden und die ABF1-TFBS in dieser Konfiguration downstream im transkribierten Bereich gelegen ist (Abschnitt 2.10). Die Expression der Gene bei ABA-Behandlung wurde, wie in Abschnitt 2.15 beschrieben, mithilfe von Genevestigator untersucht. Die Abbildung 32 zeigt einen graphischen Vergleich der Anzahl ABA-induzierter Gene in der ABF1-Downstreamgruppe und den entsprechenden Kontrollgruppen mit gleicher Gruppengröße. Die einzelnen Gene dieser Gruppen sind im Anhang (Abschnitt 7.3) aufgeführt.

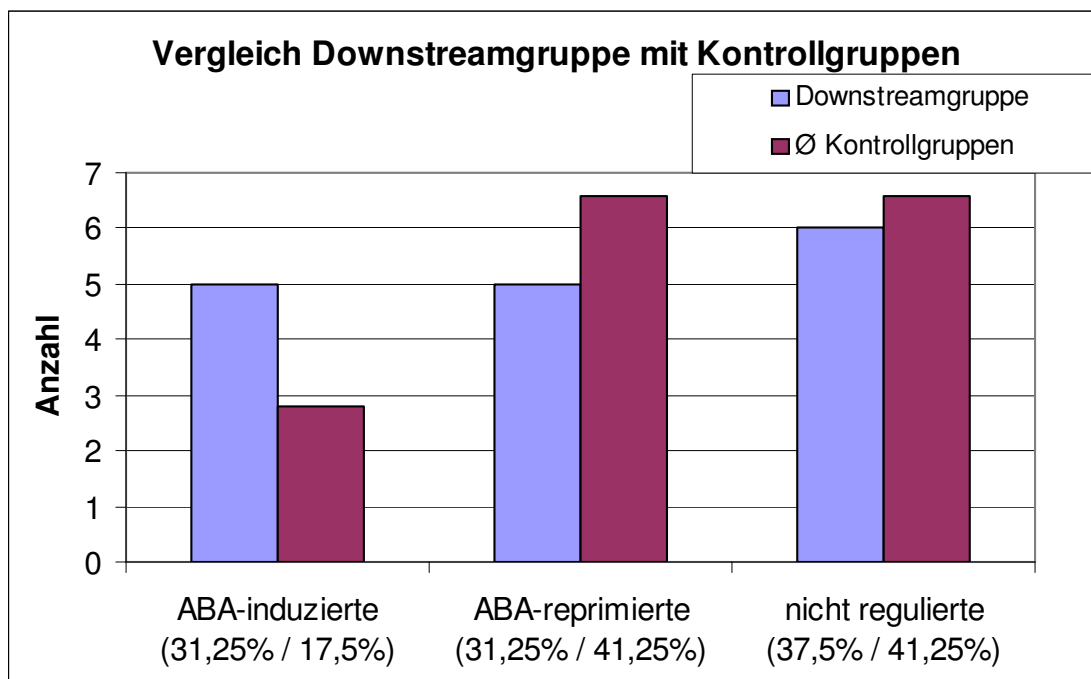


Abb. 32: Graphische Darstellung der Anteile ABA-induzierter, ABA-reprimierter und nicht ABA-regulierter Gene in der Downstreamgruppe und den entsprechenden Kontrollgruppen aus zufällig gewählten Genen. ABA-induzierte Gene treten in der ABF1\_Down-Gruppe etwas häufiger auf als in den Kontrollgruppen, jedoch ist dieser Unterschied nicht signifikant gemäß dem t-Test (Signifikanzniveau = 0,005).

### 3. Ergebnisse

Es wurde wieder mithilfe des t-Testes überprüft, inwieweit die Anzahl ABA-induzierter Gene in der ABF1-Downstreamgruppe signifikant erhöht ist gegenüber der Anzahl ABA-induzierter Gene in den Kontrollgruppen.

Parameter für den t-Test:

$N_i = 5$  (induzierte Gene in der ABF1-Downstreamgruppe)

$R_i = 2,8$  (Ø ABA-induzierte Gene in den Kontrollgruppen)

$n = 5$  (Anzahl der Gruppen aus zufällig gewählten Genen)

$s = 1,1$  (Standardabweichung)

$$t = \frac{5 - 2,8}{\frac{1,1}{\sqrt{5}}} = \frac{2,2}{0,49} = 4,47$$

Abb. 33: t-Test zum Statistischen Vergleich der Downstreamgruppe und den Kontrollgruppen

Da der kritische Wert von 5,598 unterschritten wird, ist kein signifikanter Unterschied in der Anzahl ABA-induzierter Gene feststellbar. Die Kontrollgruppen wurden mithilfe des Kolmogorov-Smirnov-Testes auf Abweichungen von einer Normalverteilung überprüft. Es konnte keine signifikante Abweichung nachgewiesen werden.

In der Gruppe mit ABF1 downstream ist der Anteil ABA-induzierter und ABA-reprimierter Gruppen vergleichbar groß bei ca. 30%. Im Gegensatz zur Prädiktionsgruppe mit ABF1-TBP-Colokalisation im upstream-Bereich ist keine Häufung ABA-induzierter Gene zu beobachten. Vielmehr ähnelt die Verteilung der Downstreamgruppe der Verteilung, die in den Kontrollgruppen vorgefunden wurde. Die Ergebnisse der oben gezeigten Untersuchungen legen den Schluss nahe, dass ABF1-TBP-Colokalisationen mit der ABF1-TFBS im upstream-Bereich von Genen mit der ABA-Induzierbarkeit dieser Gene in Zusammenhang stehen. Bei Genen, deren ABF1-TFBS im downstream-Bereich und deren TATA-Box im upstream-Bereich liegt, konnte keine signifikante Korrelation mit der ABA-Induzierbarkeit festgestellt werden.

#### 3.3.3 Bei intergenischen ABF1-TBP-Colokalisationen tritt die ABF1-Bindungsstelle bevorzugt distal zur TBP-Bindungsstelle auf

Die Reihenfolge, Distanz und Orientierung der einzelnen Bindungsstellen bei ABF1-TBP-Colokalisationen der Prädiktionsgruppe wurden im Detail untersucht. Zu diesem Zweck wurden die entsprechenden Daten tabellarisch aufgelistet (Tabelle 8) und im Hinblick auf Gemeinsamkeiten bei den ABA-induzierbaren Genen untersucht.

Dabei fällt zunächst auf, dass bei 33 von 42 Genen (78,6%) mit ABF1-TBP-Colokalisation die ABF1-Bindungsstelle upstream zu der TATA-Box liegt (Tabelle 8, Spalte „Abfolge“), während die TBP-Bindungsstelle proximal am Genstart lokalisiert ist. Die neun am stärksten ABA-induzierbaren Gene (At3g63060.1, At3g48510.1, At2g22240.1, At3g23920.1, At1g53170.1, At3g57520.1, At5g14500.1, At5g25240.1 und At5g18670.1) besitzen eine ABF1-TBP-Colokalisation in dieser Konfiguration (Tabelle 8, Spalte Expression, fettgedruckt). Von den insgesamt 21 ABA-induzierbaren Genen (Tabelle 8, Expression >1) weisen 17 Gene die ABF1-TBP-Colokalisation in dieser Konfiguration auf. Es gibt also eine bevorzugte Abfolge der beiden Bindungsstellen, die allerdings nicht nur bei den ABA-induzierbaren Genen auftritt, sondern bei allen Genen mit ABF1-TBP-Colokalisation bevorzugt zu sein scheint.

Bei sechs ABA-induzierten Genen in der Prädiktionsgruppe ist diese Regel durchbrochen, und die ABF1-Bindungsstelle liegt zwischen der TATA-Box und dem Anfang des Gens, diese sechs Gene weisen bei ABA-Behandlung Induktionsfaktoren  $\leq 1,661$  auf.

Weiterhin wurden die Abstände zwischen den ABF1- und TBP-Bindungsstellen bestimmt. Der Abstand beider Bindungsstellen ist jedoch höchstwahrscheinlich nicht das entscheidende Kriterium für ABA-Induktion, da keine Korrelation zwischen starker Induktion und einem bestimmten Abstandsbereich gefunden wurde.

### 3. Ergebnisse

Tabelle 8: Detaillierte Auflistung der einzelnen ABF1-TBP-Colokalisationen mit Sequenz. N.d: keine Expressionsänderung. **Grau** = nicht auf dem Chip repräsentiert; \* = kein ATG oder 5'UTR annotiert

AGI	Expression	Sequenz (ABF1-TBP)	Abfolge (distal nach proximal)	Abstand der TFBS	Ori ABF1 (relativ zum Gen)	Genstart
At1g03030.1	1,266	<u>gccttttaaagcgaaggagcatgacacgtgtc</u>	TBP-ABF1	22bp	Beide	TSS
At1g04820.1	n.d	<u>acacgtgtcgttatatacagatgggtacttctgtttataaataca</u>	ABF1-TBP	33bp	+	TSS
At1g07590.1	n.d	<u>gccacgtgtatggtggacataatccgaatctctctcttctctataaatagt</u>	ABF1-TBP	41bp	-	TSS
At1g10590.1	n.d	<u>tatataaaaaaataaagggtagtttaagaagatgttcagaaacatgacacgtgt</u>	TBP-ABF1	49bp	-	TSS
At1g11870.1	n.d	<u>acacgtggcaagtcacaccagttacagatccaaacgacaacgttttaaatagc</u>	ABF1-TBP	41bp	+	TSS
At1g19860.1	1,446	<u>gacacgtgtatatagg</u>	ABF1-TBP	4bp	-	TSS
At1g22510.1	1,157	<u>catattgttagccgcgtcttagcttaattctattgccacgtgt</u>	TBP-ABF1	33bp	-	TSS
At1g22600.1	0,832	<u>gtataaatccacgtggcgacacgtgt</u>	TBP-ABF1	17bp	-	ATG
At1g26670.1	1,594	<u>acacgtgtcaatataaaacgc</u>	ABF1-TBP	10bp	Beide	TSS
At1g35980.1		<u>acacgtggcacaatcctatgagttagcaactttatatatt</u>	ABF1-TBP	28bp	+	ATG
At1g48830.1	n.d.	<u>acacgtgtcatgttagtaattcgcagccgttagatcatcctttataaagt</u>	ABF1-TBP	38bp	+	TSS
At1g51400.1	n.d.	<u>acacgtggcaaaaacccatttgcataaaattca</u>	ABF1-TBP	21bp	+	TSS
At1g53165.1	n.d.	<u>gccacgtgtacagctttaactttgttctataaaaaaca</u>	ABF1-TBP	26bp	-	ATG
At1g53170.1	<b>3,886</b>	<u>acacgtggcaaaaacataagtggaataataacaaaaaccagttttatatg</u>	ABF1-TBP	39bp	+	TSS
At1g58340.1	1,655	<u>gccacgtgtagaacgatatcgcttcttacaactataaatagt</u>	ABF1-TBP	32bp	-	TSS
At1g67785.1		<u>acacgtggcatttatataca</u>	ABF1-TBP	6bp	+	ATG
At1g76570.1	1,661	<u>gcttttaaaagaactggcgaatatcattacacgtggc</u>	TBP-ABF1	28bp	+	ATG
At2g04100.1	1,225	<u>tgtatatatacaattatgtatgaatataattcacgcaacacgtgtc</u>	TBP-ABF1	37bp	+	TSS
At2g14740.1	1,233	<u>acacgtggcttgatagcgttatataaa</u>	ABF1-TBP	15bp	+	ATG
At2g22240.1	<b>6,767</b>	<u>gccacgtgtcctccctcctctgtctctcgtctatatatcca</u>	ABF1-TBP	29bp	-	TSS
At3g03250.1	n.d.	<u>gacacgtgtatattgacaattggctcagataagacgacagagatatatatatt</u>	ABF1-TBP	41bp	-	TSS
At3g04620.1	n.d.	<u>acacgtgtcacgttttagtaggacttacgagaagcgcaacgagtgctatataaatgcc</u>	ABF1-TBP	46bp	+	TSS
At3g11330.1	1,467	<u>actatttatactcagcaacacaagcttacacgtgtc</u>	TBP-ABF1	27bp	+	TSS
At3g12970.1		<u>acacgtgtctgctttagatgagagaatcaaagctgagctgtataaattgc</u>	ABF1-TBP	38bp	+	TSS
At3g18360.1	n.d.	<u>acacgtggctcactgtaacagactgaagcttataaaaaggc</u>	ABF1-TBP	28bp	+	ATG
At3g18950.1		<u>gccacgtgtcattcacccacgtcgtccgttcaccaactttttctctatatataca</u>	ABF1-TBP	45bp	-	TSS
At3g23920.1	<b>6,444</b>	<u>acacgtgtcacacgatcacctattctataaaaaaac</u>	ABF1-TBP	23bp	+	TSS
At3g24440.1	n.d.	<u>gttataaaaaaacaaaaaacaaatgacacgtgt</u>	TBP-ABF1	25bp	-	ATG
At3g26910.1	1,635	<u>gccacgtgtattttcatctctctcatataaatact</u>	ABF1-TBP	24bp	-	TSS
At3g46385.1		<u>gacacgtgtcctatatattattgg</u>	ABF1-TBP	11bp	Beide	*
At3g48510.1	<b>10,557</b>	<u>acacgtgtccagatcacagacgttataaaaacc</u>	ABF1-TBP	21bp	+	TSS

### 3. Ergebnisse

At3g48810.1	0,768	<u>gacacgtgtgtgtttctgcagtcactaacagtatttataca</u>	ABF1-TBP	29bp	-	ATG
At3g57520.1	<b>3,085</b>	<u>acacgtggcaacatcgcaacggctcttaccaaatctcctcgcatatataaagc</u>	ABF1-TBP	41bp	+	TSS
At3g63060.1	<b>11,207</b>	<u>gccacgtgtgttattctcgagcaccggaccaactgatataaatacg</u>	ABF1-TBP	34bp	-	TSS
At4g01940.1	n.d.	<u>gacacgtgtctccgttcgtttaaatggattgggtcgtcatagtaaagccatttatatc</u>	ABF1-TBP	47bp	Beide	TSS
<b>At4g12680.1</b>		<u>actataaatcctttaaccgccacgtgt</u>	<b>TBP-ABF1</b>	<b>18bp</b>	<b>-</b>	<b>ATG</b>
At4g16380.1	1,172	<u>acacgtgtcaaaaacgtggttggtgacagtgggatatgagactgtattataaaaagc</u>	ABF1-TBP	44bp	+	TSS
At5g14500.1	<b>1,995</b>	<u>acacgtggcagagtatcgacgctaagtctgatgcagagaacaactacaaatgcg</u>	ABF1-TBP	42bp	+	TSS
<b>At5g15970.1</b>		<u>acacgtggcaccacacaagtatctccatttggcctataaatata</u>	<b>ABF1-TBP</b>	<b>32bp</b>	<b>+</b>	<b>TSS</b>
At5g18670.1	<b>1,715</b>	<u>gacacgtgtaatgtccaaataggcaacgtttatccactaccttcatcttatataaaca</u>	ABF1-TBP	47bp	-	TSS
At5g25240.1	<b>1,994</b>	<u>gccacgtgtcctaaactcattcacctatatatagt</u>	ABF1-TBP	23bp	-	TSS
At5g60790.1	1,399	<u>acacgtggcacacccagaacaatcatctcctataaatacc</u>	ABF1-TBP	28bp	+	TSS

Die Werte in der Spalte „Expression“ geben an ob die Gene bei ABA-Behandlung induziert oder reprimiert vorliegen. Expression > 1: induzierte Gene. Expression < 1: reprimierte Gene. Die Expressionswerte der neun am stärksten induzierten Gene sind durch Fettdruck hervorgehoben.

#### 3.3.4 ABF1-TBP-Colokalisation definiert ein neues ABRE

Durch die Arbeit von Zhang [Zhang et al., 2005] wurde gezeigt, dass ABREs oft aus zwei G-Boxen oder einer G-Box und einem „coupling element“ (CE) bestehen. Die Einzelelemente der von Zhang et al. untersuchten kombinatorischen ABREs besitzen in der Mehrzahl eine Distanz von etwa 50bp, können im Einzelfall aber auch bis zu 150bp voneinander entfernt sein.

Im Falle der ABA-induzierten Gene der Prädiktionsgruppe wäre es möglich, dass neben den beiden analysierten TFBS für ABF1 und TBP noch ein weiteres interagierendes Element in Form einer G-Box oder eines CE an der Regulation beteiligt ist, das mit der ABF1-Bindungsstelle colokalisiert. Um das Auftreten eines klassischen ABREs im upstream-Bereich der Gene in der Prädiktionsgruppe auszuschließen, müssen potenziell interagierende Elemente detektiert werden. Dies kann mithilfe der Opaque2-Matrize durchgeführt werden, die sowohl G-Boxen als auch potenzielle CE-Sequenzen erkennt. Das Auftreten von klassischen ABREs im upstream-Bereich von Genen der Prädiktionsgruppe wurde durch eine Colokalisationsanalyse von ABF1- und Opaque2-Bindungsstellen untersucht.

Die Untersuchung zeigte, dass im upstream-Bereich der folgenden acht Gene der Prädiktionsgruppe auch klassische ABREs auftreten: At1g10590, At1g22510, At1g48830, At1g53165, At1g51400, At2g22240, At3g48510, At3g63060. Von diesen acht Genen sind bei ABA-Behandlung vier Gene induziert und vier weitere Gene nicht reguliert.

Werden diese acht Gene aus der Analyse herausgenommen, so befinden sich immer noch 27 Gene von ursprünglich 35 Genen in der Prädiktionsgruppe. Von den ABA-induzierbaren 21 Genen in dieser Gruppe bleiben nach Subtraktion der vier ABA-induzierten Gene mit klassischem ABRE noch 17 ABA-induzierbare Gene übrig. Diese 17 Gene besitzen im Bereich von 250bp um die ABF1-TFBS keine O<sub>2</sub>-Site und damit kein potenzielles kombinatorisches ABRE, trotzdem sind sie bei ABA-Behandlung induziert. Da die Gesamtgruppengröße durch die Subtraktion der Gene mit ABF1-O<sub>2</sub>-Colokalisation auf 27 Gene geschrumpft ist, wurden fünf neue Kontrollgruppen mit je 27 Genen generiert (Abschnitt 7.9). Anschließend wurde der t-Test auf einem Signifikanzniveau von 0,005 wie unter Abschnitt 3.3.2 beschrieben durchgeführt (Daten siehe Anhang, Abschnitt 7.9.1). Der Test zeigte, dass auch nach



Subtraktion aller Gene mit klassischem G-Box ABRE noch signifikant viele ABA-induzierbare Gene auftreten (Daten nicht gezeigt).

#### 3.3.5 Klassische ABREs können durch ABF1-Opaque2-Colokalisationen identifiziert werden

Mit dieser Analyse sollte überprüft werden, ob es möglich ist, klassische ABRE wie von Zhang et al. [2005] beschrieben auch mit den in AthaMap gespeicherten TFBS zu identifizieren. Zu diesem Zweck wurde eine Colokalisationsanalyse von ABF1- und Opaque2-Bindungsstellen im Abstand von 5-250bp durchgeführt. Der Mindestabstand von 5bp wurde gewählt, um potenzielle Überlagerung beider Matrizen zu verhindern. Auftretende Colokalisationen wurden jeweils in 50bp-Fenstern analysiert (siehe Tabelle 9).

Um die Anzahl der zu analysierenden Colokalisationen auf die potenziell funktionellen einzuschränken, wurden nur diejenigen Colokalisationen analysiert, bei denen die ABF1-TFBS intergenisch gelegen ist und einen Abstand von max. 100bp zum nächsten Genstart aufweisen. Dieses Kriterium basiert auf den Untersuchungen von Zhang et al. [2005], in denen gezeigt wurde, dass die distalere TFBS eines ABA-induzierbaren kombinatorischen Elementes nur in seltenen Fällen weiter als 100bp upstream vom TSS entfernt ist.

Mit den oben aufgeführten Parametern wurden klassische ABREs im upstream-Bereich von insgesamt 94 Genen detektiert (Tabelle 9, nächste Seite). Von diesen 94 Genen ist ein sehr hoher Anteil ABA-induzierbar (56 Gene), ein erheblich geringerer Teil ist ABA-reprimiert (11 Gene), und die restlichen 27 Gene sind nicht ABA-reguliert. Der Anteil der ABA-induzierbaren Gene in der Gruppe mit dem kleinsten Abstand (5-50bp, 65,7%) kommt dem Anteil der ABA-induzierbaren Gene in der Prädiktion von Zhang et al. nahe (67,5%).

### 3. Ergebnisse

Tabelle 9: Nichtredundante Gene mit ABF1-O2-Colokalisation, bei denen die ABF1-TFBS im upstream-Bereich liegt und in einen Abstand  $\leq 100$ bp zum nächsten Genstart lokalisiert. Die Aufschlüsselung der untersuchten Gene erfolgt nach ABA-induzierten, -reprimierten und nicht ABA-regulierten Genen. Die AGI-Nummern der identifizierten Gene sind im Anhang unter Abschnitt 7.4-7.8 aufgeführt. Es sind jeweils die Gene aufgeführt, die nur in einer der ABF1-O2-Colokalisationsgruppen auftreten. Gene, die in mehreren Gruppen auftreten, also folglich mehr als eine ABF1-O2-Colokalisation aufweisen, wurden aus der Expressionsanalyse entfernt, da der Expressionsstatus des Gens nicht einer von beiden Colokalisationen zugeordnet werden kann.

Gruppe	ABA-induzierte Gene	ABA-reprimierte Gene	Nicht regulierte Gene
ABF1_O2_5_50bp	23 (65,7%)	2 (5,7%)	10 (28,6%)
ABF1_O2_51_100bp	14 (58,3%)	2 (8,3%)	8 (33%)
ABF1_O2_101_150bp	8 (61,5%)	3 (23,1%)	2 (15,4%)
ABF1_O2_151-200bp	6 (54,5%)	1 (9,1%)	4 (36,4%)
ABF1_O2_201-250bp	5 (45,4%)	3 (27,3%)	3 (27,3%)
Insgesamt	56 Gene	11 Gene	27 Gene

Bisher kann nicht ausgeschlossen werden, dass allein die Nähe einer ABF1-Site zu einem TSS schon die ABA-Induzierbarkeit des benachbarten Gens vermitteln könnte. Um diese Möglichkeit zu untersuchen, wird im nächsten Schritt die ABA-Induzierbarkeit von Genen mit ABF1-TSS-Colokalisation untersucht.

#### 3.3.6 Das Auftreten einer einzelnen ABF1-Bindungsstelle ohne CE oder TATA-Box reicht nicht aus, um die ABA-Induzierbarkeit von Genen zu vermitteln

Es wurde untersucht, inwieweit eine einzelne ABF1-Bindungsstelle im Promotorbereich eines Gens bereits die ABA-Induzierbarkeit des nachgeschalteten Gens vermitteln kann. Zu diesem Zweck wurde eine Colokalisationsanalyse von ABF1 und dem TSS in einem Abstand von 0-50bp, analog zur ABF1-TBP-Analyse, durchgeführt. Es wurden nur Colokalisationen mit intergenischen, also upstream gelegenen ABF1-Bindungsstellen analysiert. Die detektierten Colokalisationen wurden dem jeweils nächstgelegenen Gen zugeordnet, wie auch in den zuvor gezeigten Analysen. Es wurden insgesamt 29 Gene mit einer ABF1-TSS-Colokalisation im Abstand von 0-50bp identifiziert, deren AGIs im Anhang aufgeführt sind (Abschnitt 7.10). Innerhalb dieser Gruppe besaßen zwei Gene auch eine ABF1-TBP-Colokalisation mit einem Abstand von 0-50bp, waren also bereits in der ursprünglichen Prädiktionsgruppe putativ ABA-regulierter Gene enthalten. Da die Untersuchung der ABF1-TSS-Colokalisationen gerade der Abgrenzung beider

### 3. Ergebnisse

Gengruppen voneinander dienen soll, wurden diese zwei Gene aus der Analyse herausgenommen. Sechzehn weitere Gene aus der ABF1-TSS-Gruppe fielen außerdem noch in eine oder mehrere ABF1-O2-Colokalisationsgruppen und wurden daher ebenfalls aus der ABF1-TSS-Colokalisationsgruppe entfernt, so dass von ursprünglich 29 Genen noch 11 Gene in der Analyse verbleiben.

Der Expressionsstatus dieser 11 Gene bei ABA-Behandlung wurde mittels Genevestigator untersucht (Abschnitt 2.15), und die Anzahl der ABA-induzierten Gene in der ABF1-TSS-Colokalisationsgruppe mit der durchschnittlichen Anzahl in fünf zufällig gewählten Kontrollgruppen verglichen (Abschnitt 2.14). Um die Kontrollgruppen und die ABF1-TSS-Gruppe vergleichbar zu halten, wurden für die Kontrollgruppen nur Gene zugelassen, die ebenfalls kein klassisches ABRE im upstream-Bereich enthalten. Abbildung 34 zeigt einen graphischen Vergleich der Anzahl ABA-induzierter Gene in der ABF1-TSS-Gruppe und den Kontrollgruppen.

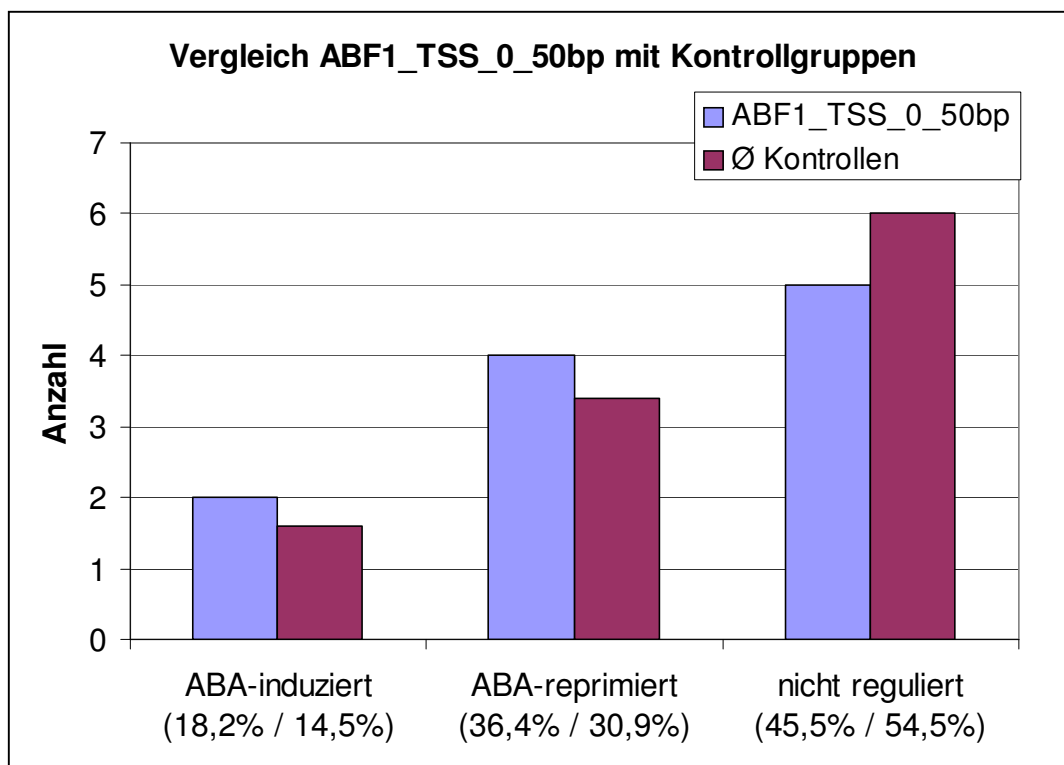


Abb. 34: Vergleich der Anzahl ABA-induzierbarer Gene in der ABF1\_TSS\_0\_50bp-Gruppe und den Kontrollgruppen. Der t-Test ( $\alpha = 0,005$ ) zeigte keinen signifikanten Unterschied zwischen beiden Gruppen.

Nachdem die Gene mit dem klassischen ABRE und einer ABF1-TBP-Colokalisation aus der Analyse entfernt wurden, ist der Anteil ABA-induzierbarer Gene in der Gruppe ABF1\_TSS\_0\_50bp nur marginal größer als der in den Kontrollgruppen bzw.

### 3. Ergebnisse

der ABF1-Downstreamgruppe. Insgesamt sind die Anteile ABA-induzierter, -reprimierter und nicht ABA-regulierter Gene in der ABF1-TSS-Gruppe und in den Kontrollgruppen vergleichbar groß.

Nun wurde mittels des t-Testes auf einem Signifikanzniveau von 0,005 überprüft, inwieweit die Anzahl ABA-induzierbarer Gene in der ABF1-TSS-Gruppe signifikant von der Anzahl in den Kontrollgruppen abweicht. Der t-Test ergab keinen signifikanten Unterschied von der Anzahl ABA-induzierbarer Genen zwischen der ABF1\_TSS\_0\_50bp-Gruppe und den Kontrollgruppen.

Parameter für t-Test:

$N_i = 2$  (induzierte Gene in der ABF1\_TSS\_0\_50bp-Gruppe)

$R_i = 1,6$  (Ø ABA-induzierte Gene in den Kontrollgruppen)

$s = 1,14$  (Standardabweichung)

$n = 5$  (Anzahl der Gruppen aus zufällig gewählten Genen)

$$t = \frac{2 - 1,6}{\frac{1,14}{\sqrt{5}}} = \frac{0,4}{0,51} = 0,78$$

Abb. 35: t-Test für den statistischen Vergleich von ABF1\_TSS\_0\_50bp-Gruppe und den entsprechenden Zufallskontrollgruppen.

Da die Colokalisation von ABF1 mit TBP anscheinend spezifisch ABA-responsive Gene selektiert und die Colokalisation von ABF1 und TSS nicht, kann dies als Indiz für eine Interaktion von ABF und TBP gewertet werden. Die reine Nähe zum TSS - ein Kriterium, das bei allen Genen der Gruppe ABF1\_TSS\_0\_50bp erfüllt ist - kann den hohen Anteil ABA-induzierbarer Gene in der Prädiktionsgruppe also nicht erklären.

#### 3.3.7 Die Gene der Prädiktionsgruppe sind auch durch osmotischen Stress und Salzstress induzierbar

Wie in der Einleitung erwähnt, ist ABA als Phytohormon für die Regulation verschiedener Stressanpassungsvorgänge zuständig, wie z.B. Kältestress, Salzstress, osmotischem Stress und Trockenstress. Daher ist anzunehmen, dass die

als ABA-induzierbar identifizierten Gene auch durch andere Stresse induziert werden können, bei denen die Signaltransduktion durch ABA erfolgt. Durch die Korrelation der Genexpressionsprofile unter ABA-Behandlung und unter verschiedenen Stressbedingungen kann möglicherweise der Pathway identifiziert werden, in den der entsprechende ABF-Regulator eingebunden ist. Um dies zu untersuchen, wurden die Genexpressionsprofile der 35 Gene der Prädiktionsgruppe bei ABA-Behandlung und bei verschiedenen Stressbehandlungen (Kältestress, Trockenstress, genotoxischer Stress, Hitzestress, osmotischer Stress, oxidativer Stress, Salzstress, UV-B-Stress, Verwundungsstress) von Genevestigator [Zimmermann et al., 2004] abgefragt und miteinander verglichen. Anschließend wurde die Korrelation der Genexpressionsprofile bei ABA-Behandlung und bei jeder einzelnen Stressbehandlung analysiert. Die Tabelle 10 zeigt das Ergebnis dieser Analyse. Die Anzahl der Gene, die sich in beiden Genexpressionsprofilen gleich verhält, wurde als „Syn“ bezeichnet. Die Anzahl der Gene, die sich bei beiden Behandlungen entgegengesetzt verhält, wurde als „Anti“ bezeichnet. Diejenigen Gene, die in einem der beiden oder beiden Genexpressionprofilen nicht exprimiert vorliegen, wurden als „Neutral“ bezeichnet. Durch Subtraktion (Syn-Anti) wurde schließlich ein Score berechnet, der die Übereinstimmung beider Genexpressionsprofile in einem einzigen Zahlenwert ausdrückt.

Tabelle 10: Korrelation der Genexpressionsprofile von Genen der Prädiktionsgruppe bei ABA-Behandlung und verschiedenen Stressbehandlungen. Im Falle von Salzstress und osmotischem Stress ist die Korrelation von ABA-Expressionsprofil und stressinduziertem Genexpressionsprofil besonders hoch.

<b>Stress</b>	<b>Syn</b>	<b>Anti</b>	<b>Neutral</b>	<b>Score =(Syn-Anti)</b>
<b>Kältestress</b>	7 (20%)	9 (25,7%)	19 (54,3%)	-2
<b>Trockenstress</b>	8 (22,9%)	6 (17,1%)	21 (60%)	2
<b>genotoxischer Stress</b>	7 (20%)	5 (14,3%)	23 (65,7%)	2
<b>Hitzestress</b>	9 (25,7%)	7 (20%)	19 (54,3%)	2
<b>hypoxischer Stress</b>	8 (22,9%)	9 (25,7%)	18 (51,4%)	-1
<b>osmotischer Stress</b>	16 (45,7%)	1 (2,9%)	18 (51,4%)	15
<b>anoxischer Stress</b>	13 (37,1%)	6 (17,1%)	16 (45,7%)	7
<b>oxidativer Stress</b>	10 (28,6%)	5 (14,3%)	20 (57,1%)	5
<b>Salzstress</b>	15 (42,9%)	2 (5,7%)	18 (51,4%)	13
<b>UV-B-Stress</b>	7 (20%)	5 (14,3%)	23 (65,7%)	2
<b>Verwundungsstress</b>	6 (17,1%)	6 (17,1%)	23 (65,7%)	0

Der Vergleich (Tab. 10) zeigte, dass die höchste Übereinstimmung zwischen dem ABA-induzierten Genexpressionsprofil und den Profilen durch osmotischen und Salzstress besteht. Eine geringe Übereinstimmung besteht zwischen dem ABA-induziertem Profil und dem Profil von Trockenstress. Bei den Profilen von ABA-Behandlung und Kältestress sind die Gegensätze am größten.

#### 3.3.8 Die Induktionsbedingungen von ABF3 korrelieren mit den Expressionsbedingungen der Gene der Prädiktionsgruppe

Die Familie der ABF-TFs (ABF1 bis ABF4; Klasse bZIP) wurde von Choi et al. detailliert untersucht [Choi et al., 2000]. Die basischen Domänen von ABF1 und ABF3 sind identisch, ebenso sind die basischen Domänen von ABF2 und ABF4 identisch. Die Expressionsprofile der vier ABF-TFs sind jedoch deutlich unterschiedlich; ABF1 ist durch Kälte oder ABA-Behandlung induzierbar und schwach induzierbar durch Salzstress [Choi et al., 2000]. ABF2 ist durch ABA-Behandlung und Salzstress stark induzierbar, und etwas schwächer induzierbar durch Kälte. ABF3 ist nur durch ABA-Behandlung und Salzstress induzierbar, und nicht induzierbar durch Kälte oder Trockenheit. ABF4 ist durch Trockenheit, Kälte, Salzstress und ABA-Behandlung annähernd gleich stark induzierbar.

Bei den Genen der Prädiktionsgruppe besteht eine hohe Korrelation der Genexpressionsprofile bei ABA-Behandlung und bei Salz- bzw. osmotischem Stress. Dies ist ein deutlicher Hinweis darauf, dass nicht ABF1 selbst, sondern ein verwandter TF den tatsächlichen Regulator der Gene in der Prädiktionsgruppe darstellt.

Da die Gene der Prädiktionsgruppe nicht kälteinduzierbar sind, kommen ABF1, ABF2 und ABF4 als Regulatoren praktisch nicht in Frage. Einzig ABF3 ist durch Salzstress induzierbar, nicht jedoch durch Kälte, und ist damit der wahrscheinlichste Regulator.

## 4. Diskussion

### 4.1 AthaMap ist die erste genomweite Karte von TFBS in *Arabidopsis thaliana*

AthaMap wurde als eine genomweite Karte von Transkriptionsfaktorbindungsstellen (TFBS) erstellt. Bei der Konzeption von AthaMap wurde auf Erweiterbarkeit, Flexibilität und die möglichst umfassende Abdeckung bekannter TF-Bindungssequenzen besonderer Wert gelegt.

Die Identifizierung von TFBS in AthaMap erfolgte mithilfe von Alignmentmatrizen, nicht mit Consensus-Sequenzen. Die verwendeten Matrizen wurden der Literatur entnommen, es wurden alle bis dato publizierten pflanzlichen Alignmentmatrizen verwendet, soweit sie nicht redundant waren, also dieselbe Bindungssequenz beschreiben.

Im Gegensatz zu AthaMap wurde in vergleichbaren Ressourcen wie z.B. AGRIS [Davuluri et al., 2003] (<http://arabidopsis.med.ohio-state.edu/AtcisDB/>), PLACE [Higo et al., 1999] (<http://www.dna.affrc.go.jp/PLACE/>) und PLANTCare (<http://intra.psb.ugent.be:8080/PlantCARE/>) [Lescot et al., 2002] auf Consensussequenzen zur Detektion potenzieller TFBS zurückgegriffen. Diese Methode ist vor allem dann vorteilhaft, wenn zu wenige Bindungsstellen aus Experimenten bekannt sind um eine brauchbare Matrize mit ausreichendem Informationsgehalt zu erstellen. Die Detektion von TFBS mithilfe von Consensussequenzen ist jedoch in jedem Falle invariabler als mit Matrizen; es ist zwar auch beim Consensus-Screening möglich, eine bestimmte Anzahl von „Mismatches“ einzustellen, die vom Suchprogramm beim Vergleich von Templatesequenz und Testsequenz toleriert werden, jedoch wird die gesuchte Sequenz nicht wie bei der Suche mit Matrizen gewichtet.

Im Gegensatz zu Consensussequenzen ist in Alignmentmatrizen eine so genannte „Coresequenz“ enthalten, die in einem Großteil der gebundenen Sequenzen enthalten ist, während die Flanken meistens etwas variabler sind und Nukleotidaustausche hier eher toleriert werden. Die Matrize spiegelt diese Präferenz einer Coresequenz durch den bindenden TF wieder, indem Nukleotidaustausche in der Coresequenz einer potenziellen Bindungsstelle einen starken Abzug vom individuellen Score der TFBS bedingen, während Austausche in der flankierenden Sequenz einen weniger starken Abzug vom Score bewirken.

Die TRANSFAC-Datenbank (<http://www.gene-regulation.com/>) ist eine der wenigen Ressourcen, die die Suche von TFBS in eigenen Sequenzen mithilfe von Matrizen ermöglicht [Matys et al., 2003]. Hier ist allerdings die Länge der zu durchsuchenden Sequenz beschränkt, genomweite Suchen können also nicht durchgeführt werden. Der genomische Kontext von TFBS ist jedoch höchstwahrscheinlich ein wichtiger Faktor bei der Spezifität der Erkennung funktioneller TFBS (Vergleiche Abschnitt 3.2, Identifizierung putativ funktioneller TATA-Boxen). Die AthaMap Datenbank ermöglicht weiterführende Untersuchungen über die Bedeutung der Lokalisierung von TFBS in transkribierten oder nichttranskribierten Bereichen, da Informationen über TFBS und über annotierte Gene gemeinsam in der Datenbank vorliegen.

Nicht nur die Lokalisierung von TFBS zur Genstruktur ist bedeutsam, sondern auch die Lokalisierung der TFBS relativ zueinander kann relevant für ihre Funktionalität sein. Dies ist der Fall bei kombinatorischen Elementen, wo zwei DNA-gebundene TFs interagieren und gemeinsam die Transkription von Zielgenen regulieren. Eine charakteristische Eigenschaft von kombinatorischen Elementen ist die räumliche Nähe der interagierenden TFBS. Beispielsweise wurde vom as-1-Element gezeigt, dass die beiden TFBS nur bis zu einem bestimmten Abstand voneinander (bis zu 15bp) synergistisch wirken [Krawczyk et al., 2002].

Die räumlich begrenzte Interaktion von zwei oder mehr TFBS bzw. ihren bindenden Transkriptionsfaktoren stellt wahrscheinlich einen wesentlichen Mechanismus zur Unterscheidung von funktionalen Bindungsstellen und zufällig auftretenden potenziellen Bindungsstellen dar. Es wurde gezeigt, dass in Säugetiergenomen die Bindungsstellen synergistischer TFs häufiger colokalisieren als die Bindungsstellen nicht synergistischer TFs [Hannenhalli & Levy 2002]. Demzufolge könnte die Analyse der Colokalisation von TFBS potenziell auch in Pflanzen Hinweise auf Synergismen zwischen TFs geben.

Aus diesem Grund wurde eine Funktion zur Durchführung von Colokalisationsanalysen in die AthaMap Datenbank implementiert. Diese Funktion wurde genutzt, um die Position kombinatorischer Elemente anhand bekannter Parameter zu detektieren (vergleiche Abschnitt 3.1.1). Ferner wurde sie zur Erstellung einer spezifischen Prädiktion über ABA-regulierte Gene eingesetzt (siehe Abschnitt 3.3). AthaMap ist zurzeit die einzige Ressource, die die Detektion von



potenziellen kombinatorischen Elementen durch genomische Colokalisationsanalysen von TFBS ermöglicht.

In der AthaMap Datenbank sind eine Auswahl bereits bekannter kombinatorischer Elemente wie das pathogeninduzierte as-1-Element, das ocs1-Element sowie die gewebespezifischen MYC/MYB- und MADS/MADS-Elemente identifiziert und annotiert. Ferner wurden die potenziellen Bindungsstellen des TFs RAV1, der zwei Bindungsdomänen besitzt, identifiziert (Abschnitt 3.1.1). Die AthaMap Datenbank ist damit die einzige Ressource, die gleichermaßen einzelne TFBS und kombinatorische Elemente enthält, und diese im genomischen Kontext darstellt.

Bei kombinatorischen Elementen kann die Bindungsspezifität für die DNA-Sequenz verändert oder herabgesetzt sein [Krawczyk et al., 2002]. Abweichungen von der Consensus-Bindungssequenz eines TFs schwächen potenziell die Protein-DNA-Bindung, auf der anderen Seite kann die Protein-Protein-Interaktion, die zwischen beiden TFs bestehen, kann die jeweilige Protein-DNA-Bindung jedes einzelnen TFs potenziell stabilisieren.

Daher ist es vorteilhaft, nicht nur hochkonservierte TFBS in AthaMap darzustellen und in Analysen miteinbeziehen zu können, sondern auch weniger stark konservierte TFBS. Um trotzdem bei Bedarf die Menge der dargestellten TFBS einschränken zu können, wurde eine Restriktionsfunktion in die AthaMap Datenbank implementiert (Abschnitt 3.1.4). Diese ermöglicht die Einschränkung der dargestellten TFBS auf höher als Threshold konservierte TFBS. Gleichermäßen wurde eine Funktion implementiert, die es in Colokalisationsanalysen ermöglicht, die Konservierung der TFBS anhand des Scores zu wählen.

#### **4.2 AthaMap enthält putativ funktionelle TATA-Boxen**

Im Rahmen dieser Arbeit sollten die putativ funktionellen TATA-Boxen im Genom von *Arabidopsis* annotiert werden. Das zunächst unter Standardbedingungen (PATSER-definierter Threshold) durchgeführte Matrizenscreening führte zur Identifizierung von über 200.000 potenziellen TATA-Boxen, also erheblich mehr als Gene im *Arabidopsis*-Genom annotiert sind [Haas et al., 2005]. Dies ist zumindest teilweise auf die Struktur der TATA-Box Matrize zurückzuführen; In A-T-reichen Bereichen des *Arabidopsis*-Genoms treten TATA-Boxen potenziell gehäuft auf. Außerdem führt die

palindromische Sequenz der TATA-Box (Consensus: TATAWATA [Shahmuradov et al., 2003]) dazu, dass eine potenzielle Bindungsstelle in *sense*-Richtung sehr häufig auch eine potenzielle TATA-Box in *antisense*-Richtung darstellt, wobei beide Bindungsstellen deckungsgleich sind. Um diesem Effekt entgegenzuwirken, wurde bei TATA-Boxen generell auf die Annotation einer Orientierung verzichtet, und positionsgleiche TBP-Bindungsstellen, die sich perfekt überlagern wie oben beschrieben, wurden nur einmal an der jeweiligen Position annotiert.

Die Eliminierung doppelt auftretender TATA-Boxen brachte bereits eine erhebliche Reduktion der potenziellen Bindungsstellen, führte jedoch noch nicht zur Identifizierung der putativ funktionellen. Eine Positionsanalyse aller detektierten potenziellen TATA-Boxen zeigte, dass diese angereichert im upstream-Bereich von Start-ATGs vorliegen. Eine detailliertere Positionsanalyse derjenigen TATA-Boxen, die im upstream-Bereich von Genen mit annotiertem TSS liegen, zeigte, dass die Position von TATA-Box und TSS miteinander korrelieren (Abschnitt 3.2.3). Es gibt in *Arabidopsis* eine deutliche Bevorzugung des Abstandes von  $35\text{bp} \pm 5\text{bp}$  zwischen TATA-Box und TSS. Insgesamt gibt es 3559 TATA-Boxen im upstream-Bereich von 2850 Genen, die im Abstand von 30-40bp mit einem TSS colokalisieren. Diese TATA-Boxen machen einen Anteil von 21,9% aller in AthaMap annotierten TATA-Boxen aus.

Der konservierte Abstand zwischen TATA-Box und TSS wurde in verschiedenen Systemen untersucht, und wird für eukaryontische Promotoren mit 30-35bp angegeben [Aso et al., 1994] [Zhu et al., 1995] [Molina & Grotewold 2005]. In den neusten Untersuchungen von Molina & Grotewold, in denen 12.749 Corepromotoren aus *Arabidopsis* untersucht wurden, wird der bevorzugte Abstand der TATA-Box mit 32bp upstream vom TSS angegeben. Die Streubreite des Abstandes geben die Autoren mit 30-45bp an, wobei nur die Region von -50bp bis -1bp der 12.749 Genpromotoren untersucht wurde und damit die Streubreite relativ zur untersuchten Region recht groß ist. Im Rahmen der vorliegenden Arbeit wurde die Lokalisierung von TATA-Boxen im upstream-Bereich von 16.319 Genen mit annotiertem TSS untersucht. In dem untersuchten Abstandsbereich von bis zu -100bp vom TSS wurde ein stark bevorzugter Abstand von 35bp festgestellt, mit einer Variationsbreite von  $\pm 5\text{bp}$ . Insofern sind die Ergebnisse der vorliegenden Arbeit mit den bereits publizierten Ergebnissen von Molina & Grotewold sehr gut

vergleichbar. In beiden Untersuchungen wurde dieselbe TATA-Box Matrize verwendet [Shahmuradov et al., 2003].

Da die Position der TATA-Box relativ zum TSS ein biologisches Kriterium für die Funktionalität darstellt, wurden positionelle Einschränkungen benutzt, um die putativ funktionellen TATA-Boxen zu identifizieren; TATA-Boxen wurden im Bereich 100bp upstream eines TSS als putativ funktionell annotiert. Der tolerierte Abstandsbereich zwischen TATA-Box und TSS wurde vom bevorzugten Abstand von  $35\text{bp} \pm 5\text{bp}$  auf einen Bereich von 100bp erweitert, da nicht davon ausgegangen werden kann, dass die annotierten TSS in allen Fällen punktgenau annotiert werden konnten und die der Annotation zugrundeliegenden cDNAs vollständig sind. Außerdem ist eine gewisse Variationsbreite des bevorzugten Abstandes von 35bp wahrscheinlich.

Da dieses Kriterium nur für die 16.319 Gene mit annotiertem TSS anwendbar ist, wurde es für die Identifizierung von TATA-Boxen im upstream-Bereich der 13.674 Gene ohne annotiertem TSS modifiziert. Bei den Genen ohne annotiertem TSS wurde die durchschnittliche Länge der 5'UTR (mit eventuell enthaltenen Introns) in *Arabidopsis* von 193bp zu dem 100bp-Annotationsfenster hinzuaddiert. Bei diesen Genen wurden TATA-Boxen also bis zu einem Abstand von -293bp upstream vom ATG annotiert.

Diese Kriterien für die positionelle Einschränkung von TATA-Boxen führten zur Annotation von 16.265 putativ funktionellen TATA-Boxen im *Arabidopsis*-Genom.

#### **4.2.1 Die Lokalisierung von ATG- und TSS-annotierten TATA-Boxen**

Nach der Annotation der TBP-Bindungsstellen wurde mittels Lokalisierungsanalyse relativ zum ATG untersucht, inwieweit sich die Lokalisierung von TATA-Boxen im upstream-Bereich von Genen mit und ohne annotierten TSS unterscheidet. Das ATG wurde hier als Bezugspunkt gewählt, weil es für alle annotierten Gene bekannt ist. Dabei zeigte sich, dass die TATA-Boxen, die anhand bekannter TSS-Positionen annotiert wurden, in einem klarer definierten Abstandsbereich zum ATG lokalisieren. Im Gegensatz dazu ist bei den TATA-Boxen, die bei Genen mit unbekanntem TSS annotiert wurden, die Streubreite der Lokalisierung größer (Abschnitt 3.2.6, Abb. 27).

Dies liegt darin begründet, dass sich die Position von TATA-Box und TSS direkt gegenseitig bedingen (Abschnitt 3.2.3) [Zhu et al., 1995]. Da die Position der TATA-Box zum TSS relativ stark konserviert ist, ist bei den Genen mit bekanntem TSS lediglich der Abstand TSS-ATG die variable Größe, die letztlich den Abstand der TATA-Box zum ATG bestimmt.

Die größere Streuungsbreite in der Lokalisierung der ATG-annotierten TATA-Boxen erklärt sich daraus, dass zwischen TATA-Box und Start-ATG nur ein indirekter Zusammenhang besteht. Einerseits gibt es eine bevorzugte 5'UTR-Länge zwischen ATG und TSS, andererseits gibt es einen bevorzugten Abstand zwischen TSS und funktioneller TATA-Box. Zusammen addiert ergeben die beiden Abstände den bevorzugten Abstand zwischen TATA-Box und ATG. Da sich die Streuungsbreite der beiden Abstände ebenfalls summiert, ist die Variabilität der TATA-Box Lokalisierung bei Genen ohne bekannten TSS größer.

Es ist jedoch beim Vergleich der Lokalisierung beider TBP-Populationen zu erkennen, dass die ATG-annotierten TBP-TFBS nicht grundsätzlich anders lokalisieren als die TSS-annotierten. Bei beiden Populationen von TATA-Boxen beträgt der bevorzugte Abstand zum ATG ca. 100bp (Abschnitt 3.2.6, Abb. 27).

Insgesamt ist bei der Population der ATG-annotierten TBP-TFBS davon auszugehen, dass aufgrund der größeren Streuungsbreite mehr falsch positive TATA-Boxen enthalten sind. Es ist jedoch gleichfalls davon auszugehen, dass die putativ funktionellen TATA-Boxen in der überwiegenden Mehrzahl der Fälle mit annotiert wurden.

#### **4.2.2 Multiple TATA-Boxen treten häufiger vor Genen ohne bekanntem TSS auf**

Es wurde analysiert, wie häufig mehrfache TATA-Boxen vor Genen mit bekanntem TSS und ohne bekannten TSS auftreten. Diese Analyse zeigte, dass mehrfache TATA-Boxen etwa doppelt so häufig vor Genen mit unbekanntem TSS (2063 Gene) auftreten als vor Genen mit bekanntem TSS (1080 Gene) (Abschnitt 3.2.6).

Das Auftreten multipler TATA-Boxen könnte in einigen Fällen auf alternative Transkriptionsstartpunkte des jeweiligen Gens zurückzuführen sein. Da diese in vielen Fällen 100-200bp weiter upstream liegen, ist das Auftreten zusätzlicher TATA-Boxen in diesem Bereich wahrscheinlich. In Abb. 27 (Abschnitt 3.2.6) ist erkennbar,

dass TATA-Boxen bis zu einem Abstand von ca. 300bp upstream des nächsten ATG gehäuft auftreten, wobei die durchschnittliche 5'UTR bei > 70% der *Arabidopsis* Gene lediglich 193bp lang oder kürzer ist. Es erscheint plausibel, dass sich in dem Bereich bis -300bp alternative TSS befinden können, deren Benutzung möglicherweise durch alternative TATA-Boxen reguliert werden könnte.

Ein Teil der auftretenden multiplen TATA-Boxen ist jedoch mit Sicherheit auf falsch positiv identifizierte TATA-Boxen zurückzuführen. Legt man den Anteil von Genen mit bekanntem TSS und TATA-Boxen (28,3%) als Basis zugrunde und geht weiterhin davon aus, dass der Anteil TATA-Box haltiger Gene mit unbekanntem TSS (41%) theoretisch etwa gleich groß sein müsste, so wären speziell in der Gengruppe ohne bekannten TSS ca. 1.750 Gene von 5.611 Genen mit potenziell falsch positiven TATA-Boxen annotiert. Bezogen auf alle Gene, für die mindestens eine TATA-Box im upstream-Bereich annotiert wurde (10.222 Gene), ist dies ein Anteil von 17,1%.

Der höhere Anteil potenziell falsch positiver TATA-Boxen bei den Genen ohne bekannten TSS liegt zum einen darin begründet, dass der Bereich für die TBP-Annotation bei den Genen ohne bekanntem TSS erheblich größer (293bp) ist als bei den Genen mit bekanntem TSS (100bp), und daher mehr zufällige TBP-Bindungsstellen auftreten müssten. Wäre allerdings das Auftreten von TATA-Boxen in den Annotationsbereichen beider Gengruppen rein zufällig bedingt, so würde man bei einem dreimal größeren Bereich auch dreimal mehr TATA-Boxen erwarten. Tatsächlich findet man trotz des erheblichen Größenunterschiedes nur ca. 10% mehr Gene mit unbekanntem TSS und mindestens einer TATA-Box im upstream-Bereich. Dies kann als Indiz gewertet werden, dass die Annotation von TATA-Boxen anhand des TSS und anhand des nächsten ATGs als annähernd gleich spezifisch zu betrachten ist.

Auf der anderen Seite ist die durchschnittliche 5'UTR-Länge von 193bp, die bei den Genen mit bekanntem TSS bestimmt und auf die Gene ohne bekannten TSS übertragen wurde, als sehr konservativ einzuschätzen. Bei den Genen mit bekanntem TSS besitzen 70% der Gene eine 5'UTR der Länge 193bp oder kürzer. Trifft die Annahme zu, dass beide Gruppen von Genen vergleichbar sind, so muss davon ausgegangen werden, dass in der Gruppe der Gene ohne bekannten TSS 70% der Gene eine 5'UTR der Länge 193bp oder kürzer besitzen und dass folglich der Bereich für die TBP-Annotation bei einem Großteil der Gene ohne bekanntem TSS etwas größer als nötig gewählt wurde.

Für die Durchführung von Colokalisationsanalysen ist es jedoch wichtig, dass möglichst alle funktionellen TATA-Boxen bei der Annotation in AthaMap tatsächlich erfasst werden.

Falsch positiv annotierte TBP-Bindungsstellen gehen in einer Colokalisationsanalyse zu Lasten der Selektivität. Mit anderen Worten, es werden bei einer hypothetischen Colokalisationsanalyse mit TBP als einem Partner einzelne Colokalisationen ausgegeben, die keine biologische Bedeutung besitzen, weil beispielsweise eine potenzielle TBP-TFBS nicht tatsächlich gebunden wird.

Fälschlich nicht annotierte TATA-Boxen gehen in einer solchen Analyse zu Lasten der Sensitivität. Dies bedeutet, dass eine biologisch funktionelle Colokalisation von zwei TFBS nicht detektiert wird, weil eine der beteiligten TFBS nicht identifiziert wurde.

Im Falle der TATA-Boxen, die in AthaMap mit den oben beschriebenen positionellen Einschränkungen annotiert wurden, ist die Selektivität als sehr hoch einzuschätzen. Es wurde gezeigt, dass ein wesentlicher Anteil der TATA-Boxen im upstream-Bereich von Genen mit bekanntem TSS tatsächlich einen Abstand von 30-40bp einhält (Abschnitt 3.2.3, Abb.24), und damit auch putativ funktionell ist [Zhu et al., 1995].

Durch Analyse mehrfach auftretender TATA-Boxen wurde weiterhin gezeigt, dass die Annotation von TATA-Boxen im upstream-Bereich von Genen mit unbekanntem TSS zwar etwas weniger spezifisch ist als im upstream-Bereich von Genen mit bekanntem TSS, jedoch immerhin in derselben Größenordnung liegt. Daher ist davon auszugehen, dass bei den TATA-Boxen der Anteil der funktionellen Bindungsstellen relativ hoch einzuschätzen ist. Um zu der vorhandenen Selektivität auch die Sensitivität von Colokalisationsanalysen zu erhalten, wurde eine konservative Annotation der putativ funktionellen TATA-Boxen durchgeführt.

Die Colokalisationsanalysen von ABF1- und TBP-Bindungsstellen, die im Abschnitt 4.3 diskutiert werden, zeigen, dass mit der vorliegenden Annotation putativ funktioneller TATA-Boxen sowohl die Sensitivität wie auch die Selektivität von Colokalisationsanalysen hoch einzuschätzen sind.

##### **4.2.3 Die Mehrzahl der *Arabidopsis*-Gene besitzt keine TATA-Box im Promotor**

Nach Annotation der putativ funktionellen TATA-Boxen wurde analysiert, wie groß der Anteil der *Arabidopsis*-Gene ist, die eine TATA-Box im Promotor aufweist. Hierbei wurde zwischen Genen mit bekanntem TSS und Genen ohne annotiertem TSS unterschieden. Die Analyse zeigte, dass von den 16.319 Genen mit bekanntem TSS 4.611 Gene (28,3%) mindestens eine TATA-Box im Promotor enthalten. Von den 13.674 Genen ohne annotierten TSS besitzen 5.611 Gene (41%) mindestens eine TATA-Box. Von den insgesamt 29.993 annotierten *Arabidopsis*-Genen besitzen damit 10.222 Gene (34,1%) mindestens eine putativ funktionelle TATA-Box.

In den neusten Untersuchungen von Molina & Grotewold [Molina & Grotewold 2005], wird der Anteil der *Arabidopsis*-Gene mit TATA-Box mit 28,8% angegeben.

Die Untersuchungen von Molina basieren auf 12.749 in TAIR annotierten Full-length-cDNAs, die von TAIR bezogen wurden. Die Untersuchungen in der vorliegenden Arbeit basieren hingegen auf den von TIGR annotierten Daten, die außer den o.g. 12.749 Fl-cDNAs noch die computergenerierten Annotationsdaten beinhalten [Haas et al., 2003] [Haas et al., 2005].

Wie unter Abschnitt 4.2.1 ausgeführt, ist die Annotation von TATA-Boxen anhand von bekannten TSS etwas spezifischer als die Annotation anhand von Start-ATGs. Der hier festgestellte Anteil der Gene mit TSS und TATA-Box (28,3%) ist annähernd gleich dem Anteil der Gene mit TATA-Box, den Molina & Grotewold feststellen (28,8%). Da Molina & Grotewold von vornherein nur Gene mit bekanntem TSS untersucht haben, ist es nicht überraschend, dass die Anteile TATA-Box haltiger Gene fast identisch sind. Die Differenz in der ersten Nachkommastelle ist möglicherweise aus dem kleineren Datensatz von Molina & Grotewold zu erklären.

### **4.3 Colokalisationsanalysen von ABF1-Bindungsstellen mit TATA- und G-Boxen geben Hinweise auf ein neuartiges ABRE**

Wie in Kapitel 4.2 beschrieben, dient die Annotation putativ funktioneller TATA-Boxen zur Identifizierung von TFs, die mit dem TBP interagieren bzw. synergistisch wirken. In der vorliegenden Arbeit wurde mittels Colokalisationsanalyse untersucht, inwieweit die TATA-Box ein Bestandteil von Abscisinsäure-Response-Elementen (ABREs) sein könnte. Als potenzielle Interaktionspartner des TBP kommen die TFs der ABF-Familie in Frage, da von verschiedenen bZIP-TFs aus unterschiedlichen Organismenreichen eine Interaktion mit dem TBP bekannt ist [Friedmann et al., 2004; Zhu et al., 2002; Ransone et al., 1993; Mikaelian et al., 1993].

Durch die Colokalisationsanalyse von ABF1-Bindungsstellen und TATA-Boxen in einem Abstand von 0-50bp wurden zunächst 42 Gene identifiziert, von denen 35 Gene auf dem ATH1-Chip (Affymetrix) repräsentiert sind. Die Untersuchung der Genexpressionsprofile zeigte, dass 60% der Gene (21 Gene von 35 Genen) mit der oben beschriebenen ABF1-TBP-Colokalisation ABA-induzierbar ist (siehe Abschnitt 3.3.1). Zwei Gene (5,7%) mit ABF1-TBP-Colokalisation (At3g48810 und At1g22600) liegen bei ABA-Einwirkung reprimiert vor, und 12 Gene sind nicht ABA-reguliert. Für die Repression der beiden reprimierten Gene könnten epigenetische Effekte oder ein bisher nicht im Detail untersuchter Repressor der ABA-Aktivität wie z.B. ABR1 verantwortlich sein [Pandey et al., 2005].

Die signifikant höhere Anzahl ABA-induzierter Gene in der Gengruppe mit ABF1-TBP-Colokalisation im Vergleich zu den Kontrollgruppen ist ein Hinweis auf einen neuartigen ABRE-Typ mit Beteiligung einer TATA-Box und einer ABF1-Bindungsstelle.

Diejenigen Gene in der Gengruppe mit ABF1-TBP-Colokalisationen, die ebenfalls klassische ABREs aus zwei G-Boxen oder einer G-Box und einem CE in ihrem putativen Promotor aufweisen, wurden aus der Analyse entfernt und die statistische Signifikanz erneut getestet. Die Anzahl ABA-induzierter Gene in der ABF1-TBP-Gruppe war immer noch signifikant hoch, womit klassische ABREs, wie von Zhang et al. beschrieben [Zhang et al., 2005], als verantwortliche regulatorische Elemente ausgeschlossen werden können. Außerdem konnte durch Colokalisationsanalysen



von ABF1-Bindungsstellen mit dem TSS ausgeschlossen werden, dass eine ABF1-Bindungsstelle in Nähe eines TSS bereits ABA-Induzierbarkeit vermitteln kann. Da die genomweite Identifizierung von ABF1-TBP-Colokalisationen spezifisch ABA-induzierbare Gene selektiert, ist es wahrscheinlich, dass die Colokalisation von ABF1-Bindungsstelle und TATA-Box im putativen Promotor eines Gens einen neuen Typ ABRE darstellt.

##### **4.3.1 Die Lokalisierung der ABF1-Bindungsstelle ist bedeutsam für die Funktionalität des neuartigen ABREs**

Bei TFBS wie auch kombinatorischen Elementen würde man intuitiv davon ausgehen, dass sie entweder upstream vom TSS eines Gens (upstream-Bereich) regulatorische Wirkung entfalten können oder, seltener, im transkribierten Bereich (downstream-Bereich; z.B. Intron) [Hong et al., 2003]. Für beide Fälle sind TFs bekannt, allerdings sind bisher keine Fälle bekannt, wo ein TF gleichermaßen im Up- und downstream-Bereich regulatorische Wirkung entfalten kann. Demzufolge wäre bei einem authentischen regulatorischen Element im Regelfall ein deutlicher Unterschied zwischen Elementen mit upstream lokalisierter Bindungsstelle und downstream lokalisierter Bindungsstelle zu erwarten.

Der Vergleich der ABF1-Downstreamgruppe (Abschnitt 3.3.2, Abb. 32) mit der ABF1-TBP-Gruppe (Abschnitt 3.3.1, Abb. 30), die die ABF1-Bindungsstelle im upstream-Bereich besitzt, zeigte den erwarteten Unterschied zwischen upstream und downstream lokalisierten Elementen. In der Gengruppe mit ABF1-TBP-Colokalisation im upstream-Bereich wurde ein signifikant erhöhter Anteil ABA-induzierbarer Gene nachgewiesen, im Gegensatz zur Downstreamgruppe, bei deren Genen die ABF1-Bindungsstelle im downstream-Bereich und die TATA-Box upstream lokalisiert ist. Die Bedeutsamkeit der Lokalisierung des neuen potenziellen ABREs ist ein weiterer Hinweis darauf, dass die Colokalisation von ABF1- und TBP-Bindungsstelle im upstream-Bereich eines Gens ein neues kombinatorisches ABRE darstellt.

##### **4.3.2 Bei intergenischen ABF1-TBP-Colokalisationen lokalisiert die ABF1-Bindungsstelle bevorzugt distal**

In über 73% der untersuchten intergenischen ABF1-TBP-Colokalisationen ist die ABF1-Bindungsstelle distal zum Genstart lokalisiert, während die TATA-Box proximal liegt.

Die Tatsache, dass diese Konfiguration in den neun am stärksten ABA-induzierbaren Genen der Prädiktionsgruppe konserviert vorliegt, ist ein Indiz dafür, dass diese Abfolge der Bindungsstellen für stärkere Induktion bei ABA-Einwirkung benötigt wird. Sie ist jedoch anscheinend nicht essenziell für ABA-Induktion, denn schwächer ABA-induzierte Gene zeigen zum Teil eine andere Konfiguration in der die TATA-Box distal und die ABF1-TFBS proximal lokalisiert ist. Auch ein ABA-reprimiertes Gen und einige nicht-regulierte Gene weisen die bevorzugte Abfolge der Bindungsstellen auf. Bei der Untersuchung des Abstandes zwischen ABF1-Bindungsstelle und TATA-Box zeigte sich, dass der durchschnittliche Abstand bei den ABA-induzierten Genen mit ca. 30bp etwas geringer war als der Abstand bei den nichtregulierten Genen mit ca. 38bp. Da jedoch auch bei den stark ABA-induzierten Genen Abstände über 40bp zwischen beiden Bindungsstellen auftreten, ist der Abstand mit hoher Wahrscheinlichkeit kein Ausschlusskriterium. Zusammengefasst legen diese Daten nahe, dass bei besonders stark ABA-induzierbaren ABF1-TBP-ABREs die ABF1-Bindungsstelle bevorzugt distal in einem Abstand von ca. 20-40bp zur TATA-Box lokalisiert.

##### **4.3.3 Das ABF1/TBP ABRE wird wahrscheinlich von ABF3 reguliert**

Die Familie der ABF-TFs wurde detailliert von Choi et al., untersucht [2000]. Die DNA-Bindungsdomänen von ABF1 und ABF3 sind identisch, die basischen DNA-Bindungsdomänen von ABF2 und ABF4 sind ebenfalls identisch. Um zu untersuchen, welcher der ABF-TFs die Zielgene mit ABF1-TBP-Colokalisation regulieren könnte, wurden die Expressionsprofile der ABF-TFs und der Gene der Prädiktionsgruppe unter verschiedenen Stressbehandlungen verglichen.

Alle ABF-Faktoren werden durch ABA-Behandlung induziert. Die Expression der einzelnen ABF-TFs wird auch durch verschiedene externe Stimuli reguliert. Der TF

ABF1 ist induzierbar durch Kälte und schwach induzierbar durch Salzstress, ABF2 ist durch Salzstress stark induzierbar, und schwächer durch Kälte. ABF3 ist durch Salzstress induzierbar, aber nicht durch Kälte oder Trockenheit, und ABF4 wird durch Trockenheit, Kälte und Salzstress annähernd gleich stark induziert [Choi et al., 2000]. In Abschnitt 3.3.7 wurde die Korrelation zwischen verschiedenen Stressbehandlungen und ABA-Behandlung bei den Genen der ABF1-TBP-Gruppe untersucht. Die Korrelation zwischen ABA-Behandlung und osmotischem Stress bzw. Salzstress ist am höchsten, während die Korrelation mit Kältestress oder Trockenheit sehr niedrig ist. Die Gene mit ABF1-TBP-Colokalisation sind also besonders stark durch Salzstress und osmotischen Stress induzierbar (Abschnitt 3.3.7), nicht jedoch durch Kältestress.

Der Vergleich mit den Induktionsbedingungen der ABF-TFs zeigt, dass lediglich ABF3 unter Salzstress und osmotischem Stress induziert vorliegt, nicht jedoch bei Kälte. Da ABF1, ABF2 und ABF4 auch kälteinduzierbar sind, kommen sie als Regulatoren der nicht-kälteinduzierten Zielgene nicht in Frage. ABF3 hingegen ist unter denselben Bedingungen induziert wie die Zielgene der ABF1-TBP-Gruppe und daher der wahrscheinlichste potenzielle Regulator.

## 5. Zusammenfassung

Pflanzen können auf Veränderungen der Umweltbedingungen durch eine Vielzahl von Anpassungen reagieren. In vielen Fällen wird diese Anpassung durch ein verändertes Muster der Genexpression hervorgerufen, an dessen Anfang die transkriptionellen Regulatoren stehen, die Transkriptionsfaktoren (TFs).

Zur Bindung an die DNA benötigen Transkriptionsfaktoren eine spezifische Erkennungssequenz, die je nach Faktor und DNA-Bindungsdomäne variiert.

Abhängig vom Grad der Konservierung der Erkennungssequenz befinden sich für jeden Transkriptionsfaktor einige Tausend bis Hunderttausende von potentiellen Bindungsstellen im Genom. Die Identifizierung von

Transkriptionsfaktorbindungsstellen (TFBS) in regulatorischen Bereichen ermöglicht es, Vorhersagen über die potentielle Regulation von Genen zu machen. Außerdem ist es möglich, durch Analyse der Lokalisierung von TFBS Rückschlüsse auf mögliche Interaktion der bindenden TFs zu ziehen.

Die potentiellen TFBS von 36 pflanzlichen TFs wurden durch genomweites Screening mit Alignmentmatrizen bei *Arabidopsis thaliana* identifiziert. Insgesamt wurden ca.  $7,5 \times 10^6$  potentielle Bindungsstellen detektiert und in der neu erstellten AthaMap Datenbank annotiert. Außerdem wurden anhand bekannter Parameter ca. 180.000 pflanzliche kombinatorische Elemente identifiziert und annotiert.

Für die AthaMap Datenbank wurde ein Webinterface implementiert

([www.AthaMap.de](http://www.AthaMap.de)), das frei zugänglich für alle Benutzer die potentiellen TFBS sowie die kombinatorischen Elemente im *Arabidopsis*-Genom visualisiert. Damit ist AthaMap die erste und zurzeit einzige Datenbank, die *cis*-regulatorische Elemente im genomischen Kontext darstellt.

Um den Konservierungsgrad der dargestellten TFBS anhand ihres Scores einstellen zu können wurde ein Webtool zur Restriktion der Darstellung in AthaMap implementiert. Basierend auf der Beobachtung daß die TFBS synergistisch wirkender oder interagierender TFs häufiger colokalisieren als die TFBS nicht synergistischer TFs wurde ein weiteres Webtool zur Detektion colokalisierender TFBS anhand benutzerdefinierter Parameter implementiert.

Eine besondere Stellung unter den TFs nimmt das TATA-Box bindende Protein ein, weil es wichtige Funktionen bei der Festlegung des Transkriptionsstartpunktes erfüllt.

Um detailliertere Untersuchungen über die Bedeutung und Interaktion des TBP durchführen zu können wurden die putativ funktionellen TATA-Boxen im *Arabidopsis*-Genom identifiziert und annotiert. Dabei zeigte sich, daß nur ca. ein Drittel aller *Arabidopsis*-Gene eine TATA-Box besitzt.

Anschließend wurden Colokalisationsanalysen von TATA-Boxen und Bindungsstellen für ein Mitglied der bZIP-TF-Familie (ABF1) durchgeführt.

Durch diese Analysen wurde ein neues potentiell kombinatorisches Abscisinsäure-Response-Element identifiziert, das aus einer TATA-Box und einer ABF1-Bindungsstelle besteht. Es wurde gezeigt, daß das Auftreten dieses neuen ABREs im Upstreambereich von Genen mit einer ABA-Induzierbarkeit korreliert. Ferner wurde nachgewiesen, daß die ABA-Induzierbarkeit dieser Gene nicht auf bereits bekannte klassische ABREs zurückzuführen ist. Durch Vergleich der Expressionsprofile von Zielgenen und der als Regulatoren in Frage kommenden TFs wurde ABF3 als wahrscheinlicher Regulator identifiziert.

6. Literaturverzeichnis

Abe H., Yamaguchi-Shinozaki K., Urao T., Iwasaki T., Hosokawa D., Shinozaki K. (1997) Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* (9): 1859-68

Aso T., Conaway J.W., Conaway R.C. (1994) Role of core promoter structure in assembly of the RNA polymerase II preinitiation complex. A common pathway for formation of preinitiation intermediates at many TATA and TATA-less promoters. *J Bio Chem* (269): 26575-83

Ausubel F.M et al., (1998) Yeast one hybrid system 13.4.1.-13.4.10

Ausubel F.M., Brent R., Kingston R.E., Moore D.D., Seldman J.G., Smith J.A., Struhl K. (1988) Current protocols in molecular biology. Green and Wiley Interscience, New York 2000

Basehoar A.D., Zanton S.J., Pugh B.F. (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* (5):699-709.

Bastola D.R., Pethe V.V., Winicov I. (1998) Alfin1, a novel zinc-finger protein in alfalfa roots that binds to promoter elements in the salt-inducible MsPRP2 gene. *Plant Mol Biol* (38): 1123-35

Baudry A., Heim M.A., Dubreucq B., Caboche M., Weisshaar B., Lepiniec L. (2004) TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* (39): 366-80

Bensmihen S., Giraudat J., Parcy F. (2005) Characterization of three homologous basic leucine zipper transcription factors (bZIP) of the ABI5 family during *Arabidopsis thaliana* embryo maturation. *J Exp Bot.* (56): 597-603

Bertrand C., Benhamed M., Li Y.F., Ayadi M., Lemonnier G., Renou J.P., Delarue M., Zhou D.X. (2005) Arabidopsis HAF2 gene encoding TATA-binding protein (TBP)-associated factor TAF1, is required to integrate light signals to regulate gene expression and growth. *J Biol Chem* (280): 1465-73

Bluthgen N., Kielbasa S.M., Herzel H. (2005) Inferring combinatorial regulation of transcription in silico. *Nucleic Acids Res* (33): 272-9

Bomblies K., Dagenais N., Weigel D. (1999) Redundant enhancers mediate transcriptional repression of AGAMOUS by APETALA2. *Dev Biol.* (216): 260-4

Burley S.K., Roeder R.G. (1996) Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem.* (65): 769-99

Cartharius K., Frech K., Grote K., Klocke B., Haltmeier M., Klingenhoff A., Frisch M., Bayerlein M., Werner T. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* (21): 2933-42

Chandrasekharan M.B., Bishop K.J., Hall T.C. (2003) Module-specific regulation of the beta-phaseolin promoter during embryogenesis. *Plant J* (33): 853-66

Chalkley GE, Verrijzer CP. (1999) DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J.* (17):4835-45

Chang C, Meyerowitz EM. (1986) Molecular cloning and DNA sequence of the Arabidopsis thaliana alcohol dehydrogenase gene. *Proc Natl Acad Sci U S A* (83): 1408-12

Chen W., Singh K.B. (1999) The auxin, hydrogen peroxide and salicylic acid induced expression of the Arabidopsis GST6 promoter is mediated in part by an ocs element. *Plant J* (19): 667-77

Choi H., Hong J., Ha J., Kang J., Kim S.Y. (2000) ABFs, a family of ABA-responsive element binding factors. *J. Biol. Chem.* (275): 1723-1730

Dai S., Petruccelli S., Ordiz M.I., Zhang Z., Chen S., Beachy R.N. (2003) Functional analysis of RF2a, a rice transcription factor. *J Biol Chem* (278): 36396-402

Davuluri R.V., Sun H., Palaniswamy S.K., Matthews N., Molina C., Kurtz M., Grotewold E. (2003) AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. *BMC Bioinformatics* (4): 25

Debeaujon I., Nesi N., Perez P., Devic M., Grandjean O., Caboche M., Lepiniec L. (2003) Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. *Plant Cell* (15): 2514-31

De Pater S., Greco V., Pham K., Memelink J., Kijne J. (1996) Characterization of a zinc-dependent transcriptional activator from Arabidopsis. *Nucleic Acids Res* (24): 4624-31

Despres C., DeLong C., Glaze S., Liu E., Fobert P.R. (2000) The *Arabidopsis* NPR1/NIM1 protein enhances the DNA binding activity of a subgroup of the TGA family of bZIP transcription factors. *Plant Cell* (2):279-90

Diaz I., Martinez M., Isabel-LaMoneda I., Rubio-Somoza I., Carbonero P. (2005) The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. *Plant J* (42): 652-62

Dong X., Braun E.L., Grotewold E. (2001) Functional conservation of plant secondary metabolic enzymes revealed by complementation of Arabidopsis flavonoid mutants with maize genes. *Plant Physiol* (127): 46-57



Dowd C., Wilson I.W., McFadden H. (2004) Gene expression profile changes in cotton root and hypocotyl tissues in response to infection with *Fusarium oxysporum* f. sp. vasinfectum. *Mol Plant Microbe Interact* (17): 654-67

Edwards D., Murray J.A., Smith A.G. (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in *Arabidopsis*. *Plant Physiol* (117): 1015-22

Egea-Cortines M., Saedler H., Sommer H. (1999) Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *EMBO J* (18): 5370-9

Ferl R. J., Laughner B.H. (1998) In vivo detection of regulatory factor binding sites of *Arabidopsis thaliana* Adh. *Plant Mol Biol* 12: 357-366

Finnegan EJ, Peacock WJ, Dennis ES. (2000) DNA methylation, a key regulator of plant development and other processes. *Curr Opin Genet Dev.* (10): 217-23

Fransz P.F., de Jong J.H. (2002) Chromatin dynamics in plants. *Curr Opin Plant Biol* 5: 560-7

Frech K., Danescu-Mayer J., Werner T. (1997) A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter. *J Mol Biol* (270): 674-87

Friedman J.S., Khanna H., Swain P.K., Denicola R., Cheng H., Mitton K.P., Weber C.H., Hicks D., Swaroop A. (2004) The minimal transactivation domain of the basic motif-leucine zipper transcription factor NRL interacts with TATA-binding protein. *J. Biol. Chem.* (279): 47233-41

Fujita Y., Fujita M., Satoh R., Maruyama K., Parvez M.M., Seki M., Hiratsu K., Ohme-Takagi M., Shinozaki K., Yamaguchi-Shinozaki K. (2005) AREB1 Is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. *Plant Cell* (17): 3470-88

Furihata T., Maruyama K., Fujita Y., Umezawa T., Yoshida R., Shinozaki K., Yamaguchi-Shinozaki K. (2006) Abscissic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. *Proc Natl Acad Sci U S A*. (6):1988-93.

Grotewold E., Drummond B.J., Bowen B., Peterson T. (1994) The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell* (76): 543-53

Grotewold E., Chamberlin M., Snook M., Siame B., Butler L., Swenson J., Maddock S., Clair G.S., Bowen B. (1998) Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. *Plant Cell* (10): 721-40

Grotewold E. (2005) Plant metabolic diversity: a regulatory perspective. *Trends Plant Sci* (10): 57-62

Gubler F., Raventos D., Keys M., Watts R., Mundy J., Jacobsen J.V. (1999) Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. *Plant J* (17): 1-9

Guilfoyle T.J., Ulmasov T., Hagen G. (1998) The ARF family of transcription factors and their role in plant hormone-responsive transcription. *Cell Mol Life Sci*. (54): 619-627

Gusmaroli G., Tonelli C., Mantovani R. (2001) Regulation of the CCAAT-Binding NF-Y subunits in Arabidopsis thaliana. *Gene* (264): 173-85

Haas B.J., Delcher A.L., Mount S.M., Wortman J.R., Smith R.K. Jr, Hannick L.I., Maiti R., Ronning C.M., Rusch D.B., Town C.D., Salzberg S.L., White O. (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* (31): 5654-66

Haas B.J., Wortman J.R., Ronning C.M., Hannick L.I., Smith R.K. Jr., Maiti R., Chan A.P., Yu C., Farzad M., Wu D., White O., Town C.D. (2005) Complete reannotation of the Arabidopsis genome: methods, tools, protocols and the final release. *BMC Biol* (1): 7.

Hannenhalli S, Levy S. (2002) Predicting transcription factor synergism. *Nucleic Acids Res* (30): 4278-84

Hartmann U., Valentine W.J., Christie J.M., Hays J., Jenkins G.I., Weisshaar B. (1998) Identification of UV/blue light-response elements in the Arabidopsis thaliana chalcone synthase promoter using a homologous protoplast transient expression system. *Plant Mol Biol* (36): 741-54

Hehl, R., N. O. Steffens, E. Wingender. (2004) Isolation and analysis of gene regulatory sequences. In: Handbook of Plant Biotechnology, Klee H, and Christou P, eds., (Wiley and Sons Ltd.), pp. 81-102.

Heine G.F., Hernandez J.M., Grotewold E. (2004) Two cysteines in plant R2R3 MYB domains participate in REDOX-dependent DNA binding. *J Biol Chem.* (36):37878-85.

Hertz G.Z., Stormo G.D. (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* (15): 563-77

Higo K., Ugawa Y., Iwamoto M., Korenaga T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* (27): 297-300

Hirayama T., Shinozaki K. (1996) A cdc5+ homolog of a higher plant, Arabidopsis thaliana. *Proc Natl Acad Sci U S A* (93): 13371-6

Hong R.L., Hamaguchi L., Busch M.A., Weigel D. (2003) Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. *Plant Cell* (15): 1296-309

Hsieh T.F., Fischer R.L. (2005) Biology of Chromatin Dynamics. *Annu Rev Plant Biol.* (56):327-51.

Huang H., Mizukami Y., Hu Y., Ma H. (1993) Isolation and characterization of the binding sequences for the product of the Arabidopsis floral homeotic gene AGAMOUS. *Nucleic Acids Res* (21): 4769-76

Huang H., Tudor M., Weiss C.A., Hu Y., Ma H. (1995)  
The Arabidopsis MADS-box gene AGL3 is widely expressed and encodes a sequence-specific DNA-binding protein. *Plant Mol Biol* (28): 549-67

Izawa T., Foster R., Chua N.H. (1993) Plant bZIP protein DNA binding specificity. *J. Mol Biol* (230): 1131-44

Jackson J.P., Lindroth A.M., Cao X., Jacobsen S.E. (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* (416): 556-60

Jakoby M., Weisshaar B., Droge-Laser W., Vicente-Carbajosa J., Tiedemann J., Kroj T., Parcy F.; bZIP Research Group. (2002) bZIP transcription factors in Arabidopsis. *Trends Plant Sci* (7): 106-11

Jang J.Y., Kim D.G., Kim Y.O., Kim J.S., Kang H. (2004) An expression analysis of a gene family encoding plasma membrane aquaporins in response to abiotic stresses in Arabidopsis thaliana. *Plant Mol Biol* (54): 713-25

Jasencakova Z., Soppe W.J., Meister A., Gernand D., Turner B.M., Schubert I. (2003) Histone modifications in Arabidopsis- high methylation of H3 lysine 9 is dispensable for constitutive heterochromatin. *Plant J* (33): 471-80

Jiao Y., Yang H., Ma L., Sun N., Yu H., Liu T., Gao Y., Gu H., Chen Z., Wada M., Gerstein M., Zhao H., Qu L.J., Deng X.W. (2003) A genome-wide analysis of blue-light regulation of Arabidopsis transcription factor gene expression during seedling development. *Plant Physiol* (133): 1480-93

Johannesson H., Wang Y., Engstrom P. (2001) DNA-binding and dimerization preferences of Arabidopsis homeodomain-leucine zipper transcription factors in vitro. *Plant Mol Biol* (45): 63-73

Kagaya Y., Ohmiya K., Hattori T. (1999) RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res* (27): 470-8

Kagaya Y., Okuda R., Ban A., Toyoshima R., Tsutsumida K., Usui H., Yamamoto A., Hattori T. (2005) Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. *Plant Cell Physiol.* (46): 300-11

Kel A.E., Gossling E., Reuter I., Cheremushkin E., Kel-Margoulis O.V., Wingender E. (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* (31): 3576-9

Kim S., Kang J.Y., Cho D.I., Park J.H., Kim S.Y. (2004) ABF2, an ABRE-binding bZIP factor, is an essential component of glucose signaling and its overexpression affects multiple stress tolerance. *Plant J* (40): 75-87

Kosugi S., Ohashi Y. (2000) Cloning and DNA-binding properties of a tobacco Ethylene-Insensitive3 (EIN3) homolog. *Nucleic Acids Res* (28): 960-7

Kosugi S., Ohashi Y. (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J* (30): 337-48

Kozaki A., Hake S., Colasanti J. (2004) The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. *Nucleic Acids Res* (32): 1710-20

Kraus R.J., Murray E. E., Wiley S. R., Zink N. M., Loritz K., Gelembiuk G. W., Mertz J. E. (2000) Experimentally determined weight matrix definitions of the initiator and TBP binding site elements of promoters. *Nucleic Acids Res.* (24): 1531-1539

Krawczyk S., Thurow C., Niggeweg R., Gatz C. (2002) Analysis of the spacing between the two palindromes of activation sequence-1 with respect to binding to different TGA factors and transcriptional activation potential. *Nucleic Acids Res* (30): 775-81

Krusell L., Rasmussen I., Gausing K. (1997) DNA binding sites recognised in vitro by a knotted class 1 homeodomain protein encoded by the hooded gene, k, in barley (*Hordeum vulgare*). *FEBS Lett.* (408): 25-9

Lago C., Clerici E., Dreni L., Horlow C., Caporali E., Colombo L., Kater M.M. (2005) The Arabidopsis TFIID factor AtTAF6 controls pollen tube growth. *Dev Biol.* (285):91-100.

Le Gourrierc J., Li Y.F., Zhou D.X. (1999) Transcriptional activation by Arabidopsis GT-1 may be through interaction with TFIIA-TBP-TATA complex. *Plant J.* (18): 663-668

Lescot M., Dehais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y., Rouze P., Rombauts S. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* (30): 325-7

Leung J., Giraudat J. (1998) Abscisinic Acid Signal Transduction. *Annu Rev Plant Physiol Plant Mol Biol* (49): 199-222

Lotan T., Ohto M., Yee K.M., West M.A., Lo R., Kwong R.W., Yamagishi K., Fischer R.L., Goldberg R.B., Harada J.J. (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* (93): 1195-205

Malagnac F., Bartee L., Bender J. (2002) An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J* (21): 6842-52

Mantovani R. (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene* (239): 15-27

Martinez-Garcia J.F., Moyano E., Alcocer M.J., Martin C. (1998) Two bZIP proteins from *Antirrhinum* flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors. *Plant J.* (13): 489-505

Martinez-Garcia J.F., Huq E., Quail P.H. (2000) Direct targeting of light signals to a promoter element-bound transcription factor. *Science* (288): 859-63

Martinez E. (2002) Multi-protein complexes in eukaryotic gene transcription. *Plant Mol Biol* (50): 925-47

Matys V., Fricke E., Geffers R., Gossling E., Haubrock M., Hehl R., Hornischer K., Karas D., Kel A.E., Kel-Margoulis O.V., Kloos D.U., Land S., Lewicki-Potapov B., Michael H., Munch R., Reuter I., Rotert S., Saxel H., Scheer M., Thiele S., Wingender E. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* (31): 374-8

Mikaelian I., Manet E., Sergeant A. (1993) The bZIP motif of the Epstein-Barr virus (EBV) transcription factor EB1 mediates a direct interaction with TBP. *C R Acad Sci III* (316): 1424-32

Molina C., Grotewold E. (2005) Genome wide analysis of Arabidopsis core promoters. *BMC Genomics* (6): 25

Nagpal P., Ellis C.M., Weber H., Ploense S.E., Barkawi L.S., Guilfoyle T.J., Hagen G., Alonso J.M., Cohen J.D., Farmer E.E., Ecker J.R., Reed J.W. (2005) Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development* (18): 4107-18.

Niu X., Helentjaris T., Bate N.J. (2002) Maize ABI4 binds coupling element1 in abscisic acid and sugar response genes. *Plant Cell* (14): 2565-2575

Nole-Wilson S., Krizek B.A. (2000) DNA binding properties of the Arabidopsis floral development protein AINTEGUMENTA. *Nucleic Acids Res* (28): 4076-82

Obayashi T., Okegawa T., Sasaki-Sekimoto Y., Shimada H., Masuda T., Asamizu E., Nakamura Y., Shibata D., Tabata S., Takamiya K., Ohta H. (2004) Distinctive features of plant organs characterized by global analysis of gene expression in Arabidopsis. *DNA Res.* (11): 11-25

Ohta M., Matsui K., Hiratsu K., Shinshi H., Ohme-Takagi M. (2001) Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell* (13): 1959-68

Orphanides G., Lagrange T., Reinberg D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev.* (10): 2657-83

Ouwerkerk P.B., Trimborn T.O., Hilliou F., Memelink J. (1999) Nuclear factors GT-1 and 3AF1 interact with multiple sequences within the promoter of the Tdc gene from Madagascar periwinkle: GT-1 is involved in UV light-induced expression. *Mol Gen Genet.* (261): 610-22

Palena C.M., Gonzalez D.H., Chan R.L. (1999) A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein Hahb-4 with DNA. *Biochem J* (341): 81-7



Pandey G.K., Grant J.J., Cheong Y.H., Kim B.G., Li L., Luan S. (2005) ABR1, an APETALA2-domain transcription factor that functions as a repressor of ABA response in arabidopsis. *Plant Physiol* (139): 1185-93

Pan S., Sehne P.C., Ferl R.J., Gurley W.B. (1999) Specific interactions with TBP and TFIIB in vitro suggest that 14-3-3 proteins may participate in the regulation of transcription when part of a DNA binding complex. *Plant Cell* (11): 1591-602

Perl-Treves R., Foley R.C., Chen W., Singh K.B. (2004) Early induction of the Arabidopsis GSTF8 promoter by specific strains of the fungal pathogen *Rhizoctonia solani*. *Mol Plant Microbe Interact* (17): 70-80

Ransone L.J., Kerr L.D., Schmitt M.J., Wamsley P., Verma I.M. (1993) The bZIP domains of Fos and Jun mediate a physical association with the TATA box-binding protein. *Gene Expr.* (3): 37-48

Reindl A, Schöffl F. (1998) Interaction between the Arabidopsis thaliana heat shock transcription factor HSF1 and the TATA binding protein TBP. *Cell Mol Life Sci.* (436): 318-322

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290: 2105-10

Riechmann J.L., Ratcliffe O.J. (2000) A genomic perspective on plant transcription factors. *Curr Opin Plant Biol* (3): 423-34

Rhee S.Y., Beavis W., Berardini T.Z., Chen G., Dixon D., Doyle A., Garcia-Hernandez M., Huala E., Lander G., Montoya M., Miller N., Mueller L.A., Mundodi S., Reiser L., Tacklind J., Weems D.C., Wu Y., Xu I., Yoo D., Yoon J., Zhang P. (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* (1):224-8.

Romero I., Fuertes A., Benito M.J., Malpica J.M., Leyva A., Paz-Ares J. (1998) More than 80R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*. *Plant J* (14): 273-84

Sainz M.B., Grotewold E., Chandler V.L. (1997) Evidence for direct activation of an anthocyanin promoter by the maize C1 protein and comparison of DNA binding by related Myb domain proteins. *Plant Cell* (9): 611-25

Scebba F., Bernacchia G., De Bastiani M., Evangelista M., Cantoni R.M., Cella R., Locci M.T., Pitto L. (2003) *Arabidopsis* MBD proteins show different binding specificities and nuclear localization. *Plant Mol Biol* (53): 715-31

Schindler U., Beckmann H., Cashmore A.R. (1992) TGA1 and G-box binding factors: two distinct classes of *Arabidopsis* leucine zipper proteins compete for the G-box-like element TGACGTGG. *Plant Cell* (4): 1309-19

Schmid C.D., Praz V., Delorenzi M., Perier R., Bucher P. (2004) The Eukaryotic Promoter Database EPD: the impact of in silico primer extension. *Nucleic Acids Res* (32): D82-5

Sessa G., Morelli G., Ruberti I. (1993) The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. *EMBO J* (12): 3507-17

Shahmuradov I.A., Gammerman A.J., Hancock J.M., Bramley P.M., Solovyev V.V. (2003) PlantProm: a database of plant promoter sequences. *Nucleic Acids Res* (31): 114-7

Shinozaki K., Yamaguchi-Shinozaki K. (1996) Molecular responses to drought and cold stress. *Curr Opin Biotechnol* (7): 161-7

Shen Q., Zhang P., Ho T.H. (1996) Modular nature of abscisic acid (ABA) response complexes: composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley. *Plant Cell* (8): 1107-19

Singh KB. (1998) Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol* (118): 1111-20

Solano R., Nieto C., Avila J., Canas L., Diaz I., Paz-Ares J. (1995) Dual DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB.Ph3) from *Petunia hybrida*. *EMBO J* (14): 1773-84

Steffens N.O., Galuschka C., Schindler M., Bulow L., Hehl R. (2004) AthaMap: an online resource for in silico transcription factor binding sites in the *Arabidopsis thaliana* genome. *Nucleic Acids Res* (32): D368-72

Steffens N.O., Galuschka C., Schindler M., Bulow L., Hehl R. (2005) AthaMap web tools for database-assisted identification of combinatorial cis-regulatory elements and the display of highly conserved transcription factor binding sites in *Arabidopsis thaliana*. *Nucleic Acids Res* (33): W397-402

Strasburger, E. (2002) Lehrbuch der Botanik, 35 Auflage, Spektrum Akademischer Verlag Heidelberg Berlin.

Sugimoto K., Takeda S., Hirochika H. (2003) Transcriptional activation mediated by binding of a plant GATA-type zinc finger protein AGP1 to the AG-motif (AGATCCAA) of the wound-inducible Myb gene NtMyb2. *Plant J* (36): 550-64

Tang W., Perry S.E. (2003) Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. *J Biol Chem* (278): 28154-9

Tilly J.J., Allen D.W., Jack T. (1998) The CArG boxes in the promoter of the *Arabidopsis* floral organ identity gene APETALA3 mediate diverse regulatory effects. *Development* (125): 1647-57

Trobner W., Ramirez L., Motte P., Hue I., Huijser P., Lonig W.E., Saedler H., Sommer H., Schwarz-Sommer Z. (1992) GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of *Antirrhinum* floral organogenesis. *EMBO J* (11): 4693-704

Uehara Y., Takahashi Y., Berberich T., Miyazaki A., Takahashi H., Matsui K., Ohme-Takagi M., Saitoh H., Terauchi R., Kusano T. (2005) Tobacco ZFT1, a transcriptional repressor with a Cys2/His2 type zinc finger motif that functions in spermine-signaling pathway. *Plant Mol Biol* (59): 435-48

Ulmasov T., Hagen G., Guilfoyle T.J. (1999) Dimerization and DNA binding of auxin response factors. *Plant J.* (19): 309-319

van Helden J. (2003) Regulatory sequence analysis tools. *Nucleic Acids Res* (31): 3593-6

Vershinin A.V., Heslop-Harrison J.S. (1998) Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Mol Biol* (36): 149-61

Vicente-Carbajosa J., Moose S.P., Parsons R.L., Schmidt R.J. (1997)  
A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc Natl Acad Sci U S A* (94): 7685-90

Yamaguchi-Shinozaki K., Shinozaki K. (2005) Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci* (10): 88-94

Yanagisawa S., Schmidt R.J. (1999) Diversity and similarity among recognition sequences of Dof transcription factors. *Plant J* (17): 209-14

Zar J.H., (1984) Biostatistical analysis, ed. Bette Kurtz (Prentice-Hall, Inc., Englewood Cliffs, N.J.)

Zemach A., Grafi G. (2003) Characterization of *Arabidopsis thaliana* methyl-CpG-binding domain (MBD) proteins. *Plant J.* (5):565-72

Zhang W., Ruan J., Ho T.H., You Y., Yu T., Quatrano R.S. (2005) Cis-regulatory element based targeted gene finding: genome-wide identification of ABA- and abiotic stress-responsive genes in *Arabidopsis thaliana*. *Bioinformatics* (21): 3074-81

Zhu Q., Dabi T., Lamb C. (1995) TATA box and initiator functions in the accurate transcription of a plant minimal promoter in vitro. *Plant Cell* (7): 1681-9

Zhu Q., Ordiz M.I., Dabi T., Beachy R.N., Lamb C. (2002) Rice TATA binding protein interacts functionally with transcription factor IIB and the RF2a bZIP transcriptional activator in an enhanced plant in vitro transcription system.  
*Plant Cell* (14): 795-803

Zhu JK. (2002) Salt and drought stress signal transduction in plants. *Annu Rev Plant Biol.* (53): 247-73

Zhu C., Perry S.E. (2005) Control of expression and autoregulation of AGL15, a member of the MADS-box family. *Plant J.* (41): 583-94

Zimmermann P., Hirsch-Hoffmann M., Hennig L., Gruissem W. (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox.  
*Plant Physiol* (136): 2621-32

## 7. Anhang

### 7.1 Herleitung der verwendeten Matrizen

#### ABF1:

Choi et al., 2000

Abb. 4 Bild Ia:

ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGC  
ACACGTGGA  
ACACGTGTC  
ACACGTGTC  
ACACGTGTC

Matrize:

A	20	0	20	0	0	0	0	0	0	1
C	0	20	0	20	0	0	0	0	0	19
G	0	0	0	0	20	0	20	17	0	0
T	0	0	0	0	0	20	0	3	0	0

Min. Score (threshold): 10,94

Max.Score: 13,16

#### bZIP910(1):

Martinez-Garcia et al., 1998

Matrize wurde aus Paper übernommen.

A	0	14	0	0	18	0	0	0	0	0	0	10
C	0	0	0	0	0	18	0	0	0	0	18	8
G	18	4	0	18	0	0	18	0	18	17	0	0
T	0	0	18	0	0	0	0	18	0	1	0	0

Min. Score (threshold): 12,86

Max.Score: 16,68

bZIP910(2):

Martinez-Garcia et al., 1998

Matrize wurde aus Paper übernommen.

A	0	1	6	2	0	1	0	0	17	0	0	0
C	0	4	1	1	1	15	0	0	0	17	0	0
G	17	10	9	2	16	1	0	17	0	0	17	0
T	0	2	1	12	0	0	17	0	0	0	0	17

Min. Score (threshold): 10,94

Max.Score: 16,08

bZIP911(1):

Martinez-Garcia et al., 1998

Matrize wurde aus Paper übernommen.

A	1	10	0	0	22	0	0	0	0	0	4	1
C	0	0	0	0	0	22	0	0	0	0	18	21
G	21	12	0	22	0	0	22	0	22	22	0	0
T	0	0	22	0	0	0	0	22	0	0	0	0

Min. Score (threshold): 14,19

Max.Score: 17,54

bZIP911(2):

Martinez-Garcia et al., 1998

Matrize wurde aus Paper übernommen.

A	0	6	0	0	10	0	0	0	0	0	8	0
C	0	0	0	0	0	10	0	0	0	0	1	10
G	10	4	0	10	0	0	10	0	10	0	1	0
T	0	0	10	0	0	0	0	10	0	10	0	0

Min. Score (threshold): 12,1

Max.Score: 15,73

TGA1:  
Schindler et al., 1992

Figure 5, Gruppe I-IV

17-1	TACGTGACGTAG
9-1r	TATCTGACGTAG
30-2	AAAATGACGTAA
9-2	GTAATGACGTAA
58-2	GTAATGACGTAT
15-1	AATGTGACGTGT
87-2	TTGATGACGTGG
43-1.1r	GGCGTGACGTGA
30-1.1	AGCCTGACGTCA
16-1r	ATCATGACGTCT
38-2	TAAGTGACGTCT
50-1r	ATCGTGACGTTA
44-1	AATATGACGTTG
43-1.2	TCCATGACGTTT
19-1r	GTGGTGACGTTA
54-1r	CTAGTGACGTTA
100-1r	CGGCTGACGTTA
45-1r	TGGGTGACGTTA
42-2r	GGGCTGACGTTA
56-2	TTGATGACGCGG
30-1.2	AGGCTGACGCTG
4-1	CCCATGACGCAT
12-1	TTGCTGACGGGA
13-1r	TCGCTGACGGTG
5-1	TTGATGACGAGA
42-1	CTAGTGACGATA
68-1	ACGATGACGATA
51-1	ACGCTGACGATA
18-2	GTAGTGACGATA
25-2	ACCGTGACGATA
25-1	TTGATGACGATA

Matrize:

A	10	6	7	12	0	0	31	0	0	7	6	18
C	4	6	8	8	0	0	0	31	0	3	3	0
G	6	6	13	11	0	31	0	0	31	2	6	7
T	11	13	3	0	31	0	0	0	0	19	16	6

Min. Score (threshold): 5,81  
Max.Score: 10,84



TGA1a:  
Izawa et al., 1993

Figure 2

4 GCCACGTGGC  
5 GACACGTGTC  
9 TACACGTGTA  
10 CGCACGTGCG  
11 CTCACGTGAG  
12 CCCACGTGGG  
13 CACACGTGTG

Figure 3

3 GTGACGTCAC  
11 CTGACGTCAG  
15 ATGACGTCAT

Figure 7

2 GCCACGTCAC

Matrix TGA1a:

A	1	3	0	11	0	0	0	0	5	1
C	5	3	8	0	11	0	0	4	1	4
G	4	1	3	0	0	11	0	7	2	5
T	1	4	0	0	0	0	11	0	3	1

Min. Score (threshold): 5,21  
Max.Score: 10,45

Opaque2:  
Izawa et al., 1993

Fig 2:

4 GCCACGTGGC  
5 GACACGTGTC  
8 TCCACGTGGA  
9 TACACGTGTA  
11 CTCACGTGAG  
12 CCCACGTGGG  
17 AACACGTGTT

Fig 3:

3 GTGACGTCAC  
11 CTGACGTCAG  
15 ATGACGTCAT

Fig 7:

2 GCCACGTCAC

O2-Matrix:

A	2	3	0	11	0	0	0	0	5	2
C	3	4	8	0	11	0	0	4	0	4
G	4	0	3	0	0	11	0	7	3	3
T	2	4	0	0	0	0	11	0	3	2

Min. Score (threshold): 4,83  
Max.Score: 10,35

PIF3(1):  
Martinez-Garcia et al., 2000

Matrize aus Paper übernommen.

A	2	5	11	6	5	11	0	27	0	0	0	0	11	9	3	11	4	4
C	5	1	2	0	3	11	27	0	27	0	0	0	2	12	9	6	19	2
G	15	10	12	17	17	5	0	0	0	27	0	27	13	6	13	0	4	9
T	5	11	2	4	2	0	0	0	0	0	27	0	1	0	2	10	0	12

Min. Score (threshold): 10,89  
Max.Score: 19,88

PIF3(2):

Martinez-Garcia et al., 2000

Matrize aus Paper übernommen.

A	15	9	10	8	1	0	32	0	0	0	0	0	9	15	11	7	4
C	2	4	2	11	26	32	0	32	0	0	0	0	10	15	13	8	11
G	9	10	13	13	5	0	0	0	32	0	32	32	11	0	8	11	6
T	6	9	7	0	0	0	0	0	0	32	0	0	2	2	0	6	11

Min. Score (threshold): 11,18

Max.Score: 18,37

DOF2:

Yanagisawa & Schmidt, 1999

Matrize aus Paper übernommen.

A	7	7	7	7	21	21	21	0	3	7	7
C	7	6	3	2	0	0	0	0	12	4	6
G	2	4	4	2	0	0	0	21	2	6	5
T	5	4	7	10	0	0	0	0	4	4	3

Min. Score (threshold): 2,72

Max.Score: 8,49

AG:

Huang et al., 1993

Matrize aus Paper übernommen.

A	21	9	10	29	0	0	31	47	52	25	17	19	7	2	22	45	40	15
C	20	3	0	8	66	65	3	2	0	0	15	8	0	0	17	4	6	10
G	6	3	1	8	0	0	6	0	1	1	11	20	57	54	5	5	9	16
T	19	51	55	21	0	1	26	17	13	40	23	19	2	10	22	12	11	25

Min. Score (threshold): 6,32

Max.Score: 14,56

AGL3:

Huang et al., 1995

Matrize aus Paper übernommen.

A	22	16	25	27	0	0	82	40	56	35	65	25	64	0	33	52	45	21
C	29	9	8	16	92	79	1	4	0	0	1	4	0	0	14	5	23	24
G	34	29	4	13	0	0	2	3	1	0	4	3	28	92	15	7	13	26
T	10	41	58	39	3	16	10	48	38	60	25	63	3	3	33	31	14	24

Min. Score (threshold): 5,79  
Max.Score: 13,32

AGL15:

Tang & Perry 2003

Matrize wurde aus Frequenzen in Fig. 1 v. Claudia Galuschka errechnet

A	5	4	16	0	1	17	10	17	13	19	8	19	0	9	26	19
C	4	2	2	32	15	4	3	0	0	1	1	0	0	5	1	3
G	3	1	4	0	0	1	2	1	0	2	4	12	31	4	1	4
T	20	25	10	0	16	10	17	14	19	10	19	1	1	14	4	6

Min. Score (threshold): 4,48  
Max.Score: 12,09

ABI41:

ABI4(1):

Niu et al., 2002

Sites aus Figure 4, Group 1:

Nr.	Seq.
34	CACCGCCCCT
17	CACCGCCCCC
38	CACCGCCCCC
13	CACCGCCCCG
79	CACCGCCCCG
65	CACCGCCCCA
22	CACCGCCGCC
24	CACCGCCGCC
31	CACCGCCGCC
28	CACCGCCGCC
4	CACCGCCACC
61	CACCGCCCGT
8	CGCCGCCCCG
42	CGCCGCCCGA
1	CGCCGCCCGT
73	CGCCGCCCAT
3	CGCCGCCCTG
60	CGCCGCCCTG
71	CGCCGCATGA
74	CGCCACCGAA
20	TGCCGCCCCG

Matrize:

A	0	12	0	0	1	0	1	1	2	4
C	20	0	21	21	0	21	20	14	13	7
G	0	9	0	0	20	0	0	5	4	6
T	1	0	0	0	0	0	0	1	2	4

Min. Score (threshold):	8,83
Max.Score:	13,83

ABI4(2):  
Niu et al., 2002

Sites aus Figure 4, Group 2:

Nr.	Seq.
5	CGGTGCGTGT
50	CGGTGCGCTG
9	CGGTGCGCAC
63	CGGTGCGGGT
37	CGGTGCACCC
51	CGGTGCACTC
48	CGGTGCAGCC
25	CGGTGCCCCC
64	CGGTGCCTCC
35	CGGTGCCTTG
41	CGGTGCCTCT
53	CGGTGCCACT
6	CGGTGCTCCC
7	CGGTGCTCCC
69	CGGTGCTCCC
19	CGGTGCTCCT
46	CGGTGCTCTT
14	CGGTGCTGCC
32	CGGTGCTGCC
55	CGGTGCTTTC
26	CGGTGCTTTC
76	CGGTGCTTCC
77	CGGTGCTTCC
80	CGGTGCCGCC
68	CGGTGCATCC
56	CGGTGGCGGT
23	CGGTGGCTTT
59	CGGTGGGGCT
62	CGGCGCCTTC
75	CGGCGCCACC
29	CGGCGCGGTC

Matrix:

A	0	0	0	0	0	0	4	2	1	0
C	31	0	0	3	0	28	10	10	18	20
G	0	31	31	0	31	3	6	8	3	2
T	0	0	0	28	0	0	11	11	9	9

Min. Score (threshold): 8,03  
Max.Score: 12,9

### ANT:

Nole-Wilson & Krizek, 2000

Matrize aus Paper übernommen.

A	8	0	31	0	21	8	15	0	4	0	0	9	34	7	3	2
C	0	32	0	34	0	6	1	3	25	34	33	11	0	1	1	10
G	24	0	2	0	13	15	1	0	0	0	0	13	0	16	29	3
T	2	2	1	0	0	5	17	31	5	0	1	1	0	10	1	19

Min. Score (threshold): 11,06

Max.Score: 18,39

### RAV1(1) und RAV1(2):

#### RAV1(1)

Kagaya Y. et al., 1999

Matrize aus Paper übernommen.

A	14	16	7	0	47	62	0	69	9	41	34	27
C	17	11	14	62	12	4	69	0	7	4	6	9
G	12	10	35	0	9	3	0	0	32	4	10	17
T	6	14	1	0	0	0	0	0	21	20	19	16

Min. Score (threshold): 5,01

Max.Score: 10,65

#### RAV1(2):

Kagaya Y. et al., 1999

Matrize aus Paper übernommen.

A	22	13	8	40	1	0	0	0	24	12	12	8
C	12	15	49	6	56	65	0	0	7	15	14	20
G	10	8	1	8	3	0	0	58	19	18	16	5
T	18	28	7	11	5	0	65	0	5	9	10	13

Min. Score (threshold): 4,98

Max.Score: 11,21

### TEIL:

Kosugi & Ohashi, 2000

Matrize aus Paper übernommen.

A	70	1	0	31	86	1	5	3
C	6	21	1	12	0	49	61	8
G	6	0	86	3	1	2	13	4
T	5	65	0	41	0	35	8	72

Min. Score (threshold): 3,81  
Max.Score: 8,38

### AtMYB15:

Romero et al., (1998)

Matrize aus Paper übernommen.

A	4	27	64	4	0	0	0	87	0	0	0	18	4
C	73	18	18	64	0	0	0	4	0	0	0	0	0
G	9	51	18	18	100	100	0	0	100	100	0	78	87
T	14	4	0	14	0	0	100	9	0	0	100	4	9

Min. Score (threshold): 11,85  
Max.Score: 17,49

### AtMYB77:

Romero et al., (1998)

Matrize aus Paper übernommen.

A	33	57	27	6	9	68	0	50	0	0	0	68	3
C	30	17	9	30	6	3	100	0	0	0	0	3	55
G	20	17	37	20	82	23	0	50	100	0	0	26	30
T	17	9	27	44	3	6	0	0	0	100	100	3	12

Min. Score (threshold): 7,25  
Max.Score: 13,11

### AtMYB84:

Romero et al., (1998)

Matrize aus Paper übernommen.

A	22	0	18	14	0	0	0	78	0	0	0	0	4
C	18	14	4	9	0	0	0	4	0	0	0	0	32
G	42	86	78	55	100	100	0	4	100	100	0	96	60
T	18	0	0	22	0	0	100	14	0	0	100	4	4

Min. Score (threshold): 12,43  
Max.Score: 17,81



CDC5:

Hirayama & Shinozaki, (1996)

Matrize aus Paper übernommen.

A		28	17	0	2	6	91	2	2	4	2	18
C		24	26	92	2	85	2	0	93	4	55	23
G		32	55	4	0	0	2	93	5	91	27	45
T		15	2	4	96	8	4	4	0	0	16	14

Min. Score (threshold): 7,47  
Max.Score: 13,93

GAMYB:

Gubler et al., 1999

Matrize aus Paper übernommen.

A		0	25	25	0	3	5	10	5
C		16	0	0	25	13	3	10	18
G		0	0	0	0	7	17	1	2
T		9	0	0	0	2	0	4	0

Min. Score (threshold): 4,64  
Max.Score: 9,62

MYB.PH3(1):

Solano et al., 1995

Matrize aus Paper übernommen.

A		22	21	27	40	33	33	0	3	0	0	0	46	1
C		3	1	1	0	0	0	46	15	0	0	0	0	13
G		4	7	7	2	8	8	0	26	46	0	0	0	2
T		11	12	6	3	4	5	0	2	0	46	46	0	7

Min. Score (threshold): 7,65  
Max.Score: 13,34

MYB.PH3(2):

Solano et al., 1995

Matrize aus Paper übernommen.

A		12	12	11	25	0	0	0	23	0	0	0	23	1
C		1	2	2	0	0	0	0	0	0	0	0	0	5
G		6	7	8	0	27	3	0	0	27	0	0	4	4
T		6	4	4	1	0	24	27	4	0	27	27	0	1

Min. Score (threshold): 8  
Max.Score: 12,96

P:

Grotewold et al., 1994

Matrize aus Paper übernommen.

A		25	2	0	17	36	0	1	11	6
C		3	28	31	0	0	36	34	13	8
G		4	3	0	0	0	0	0	4	15
T		4	3	5	19	0	0	1	8	7

Min. Score (threshold): 4,94  
Max.Score: 10,19

GT1:

Ouwerkerk et al., 1999

Matrize aus Paper übernommen.

A		9	7	10	5	4	0	1	11	17	14	11	9
C		1	1	2	3	2	0	0	0	0	1	0	2
G		7	4	4	3	9	18	0	1	0	0	0	0
T		1	6	2	7	3	0	17	6	1	3	7	7

Min. Score (threshold): 2,72  
Max.Score: 8,94

### PCF2:

Kosugi, & Ohashi, 2002

Matrize aus Paper übernommen.

A		15	16	3	3	17	2	0	0	64	1
C		14	14	1	6	19	71	73	75	1	66
G		11	30	66	65	25	2	3	0	8	2
T		24	14	6	2	15	1	0	1	3	7

Min. Score (threshold): 6,53  
Max.Score: 12,22

### PCF5:

Kosugi, & Ohashi, 2002

Matrize aus Paper übernommen.

A		6	4	0	1	5	0	0	4	17	15
C		3	20	0	0	22	83	83	74	24	28
G		63	5	82	82	13	0	0	3	25	12
T		3	52	1	0	43	0	0	2	15	26

Min. Score (threshold): 7,48  
Max.Score: 12,26

### HVH21:

Krusell et al., 1997

Matrize aus Paper übernommen.

A		3	14	5	4	0	1	37	1	15	4	9	7
C		12	11	12	14	3	1	0	35	8	8	6	4
G		11	9	7	16	1	35	0	0	13	18	13	11
T		11	3	13	3	33	0	0	1	1	7	9	15

Min. Score (threshold): 3,82  
Max.Score: 10,96

### ALFIN1:

Bastola et al. 1998

Sequenzen aus Figure 6

(jeweils 3 Nukleotide. vor und 4 Nuk. hinter der Kernsequenz):

	1	2	3	4	5	6	7	8	9	10
R6	G	C	G	G	T	G	G	C	C	C
R7	G	G	G	G	T	G	G	G	G	A
R15	A	G	G	G	T	G	T	G	G	G
R16	G	A	G	G	T	G	G	A	G	G
R22	A	A	C	G	T	G	G	A	G	A
R25	G	G	A	G	T	G	T	G	G	T
R21	A	G	C	G	T	G	T	T	G	C
R5	T	A	G	G	T	G	G	G	A	C
M-221	-	G	T	G	T	G	G	G	G	C
Ohne1	A	A	A	G	T	G	G	G	G	G
M-218	C	A	T	G	T	G	T	G	T	G
M-187	C	A	A	G	T	G	G	T	G	C
Ohne2	G	C	G	G	T	G	C	T	G	-

### Matrize:

A	4	6	3	0	0	0	0	2	1	2
C	2	2	2	0	0	0	1	1	1	5
G	5	5	6	13	0	13	8	7	10	4
T	1	0	2	0	13	0	4	3	1	1

Min. Score (threshold): 4,16  
Max.Score: 11,23

### ATHB1:

Sessa et al. 1993

Nur Positionen von -4 bis +4

A	3	22	25	0	0	25	0	0	3
C	13	1	0	0	5	0	0	0	1
G	4	0	0	0	0	0	0	0	12
T	3	1	0	25	20	0	25	25	0

Min. Score (threshold): 5,77  
Max.Score: 9,98

**ATHB5:**

Johannesson et al., 2001

Alle Sequenzen aus Fig. 2:

```
1   tatctaattcattg
2   gtgccaatcattg
3   taccgaattattg
4   -gggtaattattg
5   cagccaatgattg
6   ----caatcattg
7   -gttcaattattg
8   --tgcaattattg
9   cccagcattattg
10  gctccaattattg
11  tgcccaataattg
12  cgggcaataattg
13  ggctctattattg
14  -tgccaatcattg
15  cctccaatgattg
16  ggtccaattattg
17  gggccaatgattg
18  ggactaatgattg
19  --ctcaatcattg
20  ggtgcaattattg
21  tgttcaattattg
22  ccaccaattattg
23  cgggtgaattattg
24  gttcgcatgattg
25  tccgcaatgattg
26  tgtgccattattg
27  tgttcaattattg
28  -ccgcaatcattg
29  tatcccattattg
30  gcctgcattattg
31  gctgcaatgattg
32  ggcacaattattg
33  ggtccaattattg
34  ccggccattattg
35  cgcccaattattg
36  ----caatcattt
37  ----caattattg
38  ctcccaattattg
39  tgcgcaatcattg
```

Alle 3' überstehenden Nukleotide (bei Sequenz Nr. 6, 9, 36, 37) wurden für die Generierung der Matrize entfernt.

Matrize wurde aus den Sequenzen in Figuren 2 im Paper neu generiert, da die Herleitung der Matrize im Paper nicht nachvollziehbar war.

### Matrize:

A	0	4	2	2	0	32	39	0	2	39	0	0	0
C	9	9	12	17	32	6	0	0	8	0	0	0	0
G	12	17	8	10	4	0	0	0	7	0	0	0	38
T	9	4	14	7	3	1	0	39	22	0	39	39	1

Min. Score (threshold): 7,99  
Max.Score: 13,59

### ATHB9:

Sessa et al., 1998

Matrize aus Paper übernommen.

A	4	5	5	3	0	0	25	26	0	0	26	0	0	17	0	5	2	0	0
C	2	3	4	8	1	6	0	0	0	0	0	0	0	0	25	1	6	3	5
G	1	3	2	2	25	0	1	0	0	26	0	0	3	9	1	4	0	4	3
T	2	3	5	5	0	20	0	0	26	0	0	26	23	0	0	7	4	3	0

Min. Score (threshold): 10,8  
Max.Score: 17,83

### HAHB4:

Palena et al. 1999

Matrize aus Paper übernommen.

A	1	25	25	0	7	25	0	3	6
C	0	0	0	0	0	0	0	1	5
G	0	0	0	0	14	0	0	3	14
T	24	0	0	25	4	0	25	18	0

Min. Score (threshold): 5,32  
Max.Score: 9,64

# AGP1:

Sugimoto et al. 2003

Figure 5, a) Kernsequenz(groß) -2 und +1

TCAGATCTAA  
ACAGATCTAG  
ACAGATCTAG  
TTAGATCTAG  
CAAGATCTAA  
CGAGATCTGA  
TGAGATCTTC  
CGAGATCTAC  
GTAGATCTAT  
ATAGATCTAT  
ACAGATCTAT  
CAAGATCTAC  
TCAGATCTGT  
TCAGATCTAT  
TCAGATCTAC  
TCAGATCCTG  
CTAGATCCGG  
TCAGATCCAC  
ACAGATCCAA

## Matrize:

A	5	2	19	0	19	0	0	0	14	4
C	5	10	0	0	0	0	19	4	0	5
G	1	3	0	19	0	0	0	0	3	5
T	8	4	0	0	0	19	0	15	2	5

Min. Score (threshold): 5,45  
Max.Score: 10,11

ZAP1:

de Pater et al., 1996

Matrize aus Paper konstruiert, Figure 5, nur erste und zweite Gruppe (mit affinity > 10)

65	TTGACCGAG
59	TTGACCGAG
46	TTGACCGAG
35	TTGACCGAG
94	TTGACCGAG
5	TTGACCGAG
100	TTGACCGAG
78	TTGACCGAG
60	TTGACCGAG
15	TTGACCTCA
66	TTGACCGAT
82	TTGACCGAG
12	TTGACCGAG
113	TTGACCGAG
110	TTGACCGAG
115	TTGACCGAG
53	TTGACTATT
48	TTGACCGAG
57	TTGACCGGT
98	TTGACCGAG
83	TTGACCGAG
41	TTGACCGAC
1	TTGACCGAG
69	TTGACCGTT
72	TTGACCGAG
92	TTGACCGAG
111	TTGACCGAG
71	TTGACCGAG
26	TTGACCGCG
61	TTGACCGAG
64	TTGACCCGC
141	TTGACCGAG
32	TTGACCGAG
144	TTGACCGAG
68	TTGACCCGC
97	TTGACCGAG
81	TTGACCTTG
140	TTGACCGAG
143	TTGACCGCG
131	TTGACCGAG
112	TTGACCGCC



52 TTGACCCCA  
 104 TTGACCGAG  
 114 TTGACCGCG  
 105 TTGACCGAG

Matrize:

A	0	0	0	45	0	0	1	33	2
C	0	0	0	0	45	44	3	6	4
G	0	0	45	0	0	0	39	3	35
T	45	45	0	0	0	1	2	3	4

Min. Score (threshold): 8,48  
 Max.Score: 12,26

ID1:  
 Kosaki et al. 2004

Figure1A (außer Sequenz Nr. 40):

5	TTTGTCGTTTT
20	TTTGTCGCTTT
11	TTTGTCGTTAT
48	TTTGTCGTATT
3	TTTGTCCTATT
6	TTTGTCGCTTA
12	CTTGTCCTTAT
18	CTCGTCGTTTT
36	TTTGCGGTTAT
29	TTTGTCGCCATA
9	TTTGTCCTTAC
4	TTTGTCCTCATT
8	CTTGTCCTTTAT
10	TTTGTTGTCCT
27	TTTTTCCCCTA

Matrize:

A	0	0	0	0	0	0	0	0	4	5	3
C	3	0	1	0	0	14	5	6	2	1	1
G	0	0	0	14	1	0	8	0	0	0	0
T	12	15	14	1	14	1	2	9	9	9	11

Min. Score (threshold): 5,71  
 Max.Score: 11,02

TBP:

Shahmuradov et al., 2003

Die vorliegende Matrize wurde aus einer Frequenzmatrize im Paper errechnet.

A	48	27	5	162	0	171	106	166	65	125	22	51
C	46	108	2	0	7	0	0	0	2	14	72	72
G	29	9	0	0	0	0	0	3	0	17	48	27
T	48	27	164	9	164	0	65	2	104	15	31	19

Min. Score (threshold): 5,16

Max.Score: 10,98

CBF:

Shahmuradov et al., 2003

Die vorliegende Matrize wurde aus einer Frequenzmatrize im Paper errechnet.

A	41	45	35	39	41	0	131	131	0	37	42	38	52
C	25	22	21	24	45	131	0	0	0	26	26	33	22
G	26	26	35	28	20	0	0	0	0	26	24	20	20
T	39	38	39	41	26	0	0	0	131	42	39	41	37

Min. Score (threshold): 3,98

Max.Score: 7,24

## 7.2 Vergleichsgruppen zufällig ausgewählter Gene zum Vergleich mit der ABF1-TBP-Prädiktionsgruppe (Gruppengröße 35 Gene)

ABA\_Random1:

Treatment			
Hormone: ABA (+)			
Probeset		AGI/Links/Annotation	
265529_at	0.599	AT2G06230	[MIPS, Tair, Tigr]hypothetical protein]
256699_at	0.513	AT3G20690	[MIPS, Tair, Tigr]F-box protein-related contains weak hit to TIGRFAM TIGR01640 : F-box protein inte...
265095_at	0.128	AT1G03880	[MIPS, Tair, Tigr]12S seed storage protein (CRB) identical to 12S seed storage protein, gi 808937 ...
260139_at	1.067	AT1G66380	[MIPS, Tair, Tigr]myb family transcription factor (MYB114) similar to myb-related protein An2 Gl:7...
251416_at	1.060	AT3G60400	[MIPS, Tair, Tigr]mitochondrial transcription termination factor-related / mTERF-related contains ...
245401_at	1.081	AT4G17670	[MIPS, Tair, Tigr]senescence-associated protein-related similar to senescence-associated protein S...
267028_at	1.050	AT2G38470	[MIPS, Tair, Tigr]WRKY family transcription factor contains Pfam profile: PF03106 WRKY DNA-bindin...
248103_at	1.041	AT5G55160	[MIPS, Tair, Tigr]small ubiquitin-like modifier 2 (SUMO) similar to ubiquitin-like protein SMT3 SP...
266181_at	1.130	AT2G02390	[MIPS, Tair, Tigr]glutathione S-transferase zeta 1 (GSTZ1) (GST18) identical to SP Q9ZVQ3 GTZ1_ARA...
! 257057_at	0.982	AT3G15310	[MIPS, Tair, Tigr][AT3G15310, expressed protein]
		AT5G32621	[MIPS, Tair, Tigr][AT5G32621, expressed protein contains Pfam profile PF04827: Protein of unknow...
254402_at	0.978	AT4G21310	[MIPS, Tair, Tigr]expressed protein predicted protein, Arabidopsis thaliana]
254128_at	0.971	AT4G24560	[MIPS, Tair, Tigr]ubiquitin-specific protease 16, putative (UBP16) similar to ubiquitin-specific p...
246102_at	0.973	AT5G32590	[MIPS, Tair, Tigr]myosin heavy chain-related similar to Myosin heavy chain, non-muscle (Zipper pro...
249911_at	0.948	AT5G22740	[MIPS, Tair, Tigr]glycosyl transferase family 2 protein similar to beta-(1-3)-glucosyl transferase...
! 256893_x_at	1.007	AT3G30660	[MIPS, Tair, Tigr]hypothetical protein]
251689_at	1.008	AT3G56500	[MIPS, Tair, Tigr]serine-rich protein-related contains some similarity to serine-rich proteins]
265910_at	1.001	AT2G25610	[MIPS, Tair, Tigr]H+-transporting two-sector ATPase, C subunit family protein similar to SP P23968...
258027_at	0.843	AT3G19515	[MIPS, Tair, Tigr]expressed protein]
249623_at	0.855	AT5G37560	[MIPS, Tair, Tigr]zinc finger protein-related contains weak similarity to zinc fingers and Pfam:PF...
263459_at	0.826	AT2G31820	[MIPS, Tair, Tigr]ankyrin repeat family protein contains ankyrin repeat domains, Pfam:PF00023]
254703_at	0.877	AT4G17960	[MIPS, Tair, Tigr]expressed protein]
! 264567_s_at	0.769	AT1G05240	[MIPS, Tair, Tigr][AT1G05240, peroxidase, putative similar to peroxidase; peroxidase ATP11a [Ara...
		AT1G05250	[MIPS, Tair, Tigr][AT1G05250, peroxidase, putative similar to peroxidase; peroxidase ATP11a [Arab...
266670_at	1.232	AT2G29740	[MIPS, Tair, Tigr]UDP-glucuronosyl/UDP-glucosyl transferase family protein contains Pfam profile: ...
252830_at	1.192	AT4G39850	[MIPS, Tair, Tigr]peroxisomal ABC transporter (PXA1) identical to peroxisomal ABC transporter PXA1...
246463_at	1.275	AT5G16970	[MIPS, Tair, Tigr]NADP-dependent oxidoreductase, putative (P1) identical to probable NADP-dependen...
251838_at	1.651	AT3G54940	[MIPS, Tair, Tigr]cysteine proteinase, putative contains similarity to cysteine proteinase Gl:4790...
246453_at	1.630	AT5G16830	[MIPS, Tair, Tigr]syntaxin 21 (SYP21) / PEP12 homolog identical to Syntaxin homolog (PEP12 homolog...
255930_at	1.723	AT1G12660	[MIPS, Tair, Tigr]hypothetical protein]
263960_at	1.829	AT2G36200	[MIPS, Tair, Tigr]kinesin motor protein-related]
258744_at	1.788	AT3G05830	[MIPS, Tair, Tigr]expressed protein]
247279_at	1.875	AT5G64310	[MIPS, Tair, Tigr]arabinogalactan-protein (AGP1) identical to gi:3883120 gb:AAC77823]
262739_at	1.532	AT1G28650	[MIPS, Tair, Tigr]lipase, putative strong similarity to lipase [Arabidopsis thaliana] Gl:1145627]
251361_at	1.545	AT3G61230	[MIPS, Tair, Tigr]LIM domain-containing protein similar to SP P29675 Pollen specific protein SF3 {...
255246_at	1.425	AT4G05640	[MIPS, Tair, Tigr]hypothetical protein]
266816_at	2.009	AT2G44970	[MIPS, Tair, Tigr]lipase-related contains weak similarity to lipase [Pseudomonas aeruginosa] gi 35...
Hormone: ABA (-)			
		Control	

13 Gene ABA-induziert  
7 Gene ABA-reprimiert  
15 Gene nicht reguliert

## ABA\_Random2:

Treatment		Hormone: ABA (+)		AGI/Links/Annotation	
Probeset					
267424_at	1.094	AT2G34800	[MIPS, Tair, Tigr]	hypothetical protein]	
256391_at	1.101	AT3G06090	[MIPS, Tair, Tigr]	hypothetical protein]	
247311_at	1.153	AT5G63960	[MIPS, Tair, Tigr]	DNA-directed DNA polymerase delta catalytic subunit, putative (POLD1) similar to...	
263380_at	1.214	AT2G40200	[MIPS, Tair, Tigr]	basic helix-loop-helix (bHLH) family protein contains Pfam profile: PF00010 heli...	
250859_at	1.218	AT5G04660	[MIPS, Tair, Tigr]	cytochrome P450, putative cytochrome P450 77A3p, Glycine max., PIR:T05948]	
253789_at	1.205	AT4G28570	[MIPS, Tair, Tigr]	alcohol oxidase-related low similarity to long chain fatty alcohol oxidase from ...	
256229_at	1.364	AT1G58200	[MIPS, Tair, Tigr]	mechanosensitive ion channel domain-containing protein / MS ion channel domain-c...	
! 263878_s_at	1.016	AT2G22040	[MIPS, Tair, Tigr]	transducin family protein / WD-40 repeat family protein similar to Pop3 (GI:3434...	
247813_at	1.020	AT5G58330	[MIPS, Tair, Tigr]	malate dehydrogenase [NADP], chloroplast, putative strong similarity to chlorop...	
253877_at	1.023	AT4G27435	[MIPS, Tair, Tigr]	expressed protein]	
248071_at	1.030	AT5G55670	[MIPS, Tair, Tigr]	RNA recognition motif (RRM)-containing protein]	
260279_at	1.059	AT1G80420	[MIPS, Tair, Tigr]	DNA repair protein, putative (XRCC1) identical to putative DNA repair protein XR...	
258499_at	0.958	AT3G02540	[MIPS, Tair, Tigr]	ubiquitin family protein contains Pfam profiles PF00240: Ubiquitin family, PF006...	
257254_at	0.963	AT3G21950	[MIPS, Tair, Tigr]	S-adenosyl-L-methionine:carboxyl methyltransferase family protein similar to SAM...	
265398_at	0.954	AT2G11000	[MIPS, Tair, Tigr]	natC N(alpha)-terminal acetyltransferase, Mak10 subunit family protein contains ...	
259359_at	0.976	AT1G13460	[MIPS, Tair, Tigr]	serine/threonine protein phosphatase 2A (PP2A) regulatory subunit B', putative s...	
246154_at	0.973	AT5G19940	[MIPS, Tair, Tigr]	plastid-lipid associated protein PAP-related / fibrillin-related weak hit to Pfa...	
251912_at	0.980	AT3G53840	[MIPS, Tair, Tigr]	protein kinase family protein contains protein kinase domain, Pfam:PF00069]	
265450_at	0.929	AT2G46620	[MIPS, Tair, Tigr]	AAA-type ATPase family protein contains Pfam profile: ATPase family PF00004]	
258435_at	0.940	AT3G16740	[MIPS, Tair, Tigr]	F-box family protein contains F-box domain Pfam:PF00646]	
255950_at	0.827	AT1G22110	[MIPS, Tair, Tigr]	expressed protein]	
255686_at	0.826	AT4G00630	[MIPS, Tair, Tigr]	K+ efflux antiporter, putative (KEA2) Monovalent cation:proton antiporter family...	
259771_at	0.835	AT1G29470	[MIPS, Tair, Tigr]	dehydration-responsive protein-related similar to early-responsive to dehydratio...	
252352_at	0.848	AT3G48185	[MIPS, Tair, Tigr]	expressed protein]	
250514_at	0.843	AT5G09550	[MIPS, Tair, Tigr]	Rab GDP dissociation inhibitor, putative strong similarity to GDP dissociation i...	
! 261974_at	0.887	AT1G64630	[MIPS, Tair, Tigr]	[AT1G64630, protein kinase family protein contains eukaryotic protein kinase do...	
		AT1G64640	[MIPS, Tair, Tigr]	[AT1G64640, plastocyanin-like domain-containing protein contains InterPro:IPR0...	
252305_at	0.775	AT3G49240	[MIPS, Tair, Tigr]	pentatricopeptide (PPR) repeat-containing protein contains Pfam profile PF01535...	
258951_at	0.678	AT3G01380	[MIPS, Tair, Tigr]	phosphatidylinositolglycan class N (PIG-N) family protein similar to phosphatidy...	
255467_at	0.590	AT4G03010	[MIPS, Tair, Tigr]	leucine-rich repeat family protein contains leucine rich-repeat domains Pfam:PF0...	
264619_at	1.551	AT2G17760	[MIPS, Tair, Tigr]	aspartyl protease family protein contains Eukaryotic and viral aspartyl protease...	
257458_at	1.642	AT2G05400	[MIPS, Tair, Tigr]	meprin and TRAF homology domain-containing protein / MATH domain-containing prot...	
260225_at	0.221	AT1G74590	[MIPS, Tair, Tigr]	glutathione S-transferase, putative similar to putative glutathione S-transferas...	
263222_at	2.352	AT1G30640	[MIPS, Tair, Tigr]	protein kinase, putative contains protein kinase domain, Pfam:PF00069]	
248901_at	2.470	AT5G46410	[MIPS, Tair, Tigr]	NLI interacting factor (NIF) family protein contains Pfam profile PF03031: NLI i...	
266778_at	7.128	AT2G29090	[MIPS, Tair, Tigr]	cytochrome P450 family protein similar to Cytochrome P450 88A3 (SP:O23051) [Arab...	
		Hormone: ABA (-)			
Control					

10 Gene ABA-induziert  
 9 Gene ABA-reprimiert  
 16 Gene nicht reguliert

## ABA\_Random3:

Treatment		Hormone: ABA (+)		AGI/Links/Annotation
Probeset				
252343_at	0.702			AT3G48610 [MIPS, Tair, Tigr]phosphoesterase family protein low similarity to SP P95246 Phospholipase C 2 pre...
246098_at	0.668			AT5G20400 [MIPS, Tair, Tigr]oxidoreductase, 2OG-Fe(II) oxygenase family protein similar to naringenin,2-oxog...
266977_at	0.774			AT2G39420 [MIPS, Tair, Tigr]esterase/lipase/thioesterase family protein similar to monoglyceride lipase from...
255513_at	0.760			AT4G02060 [MIPS, Tair, Tigr]proliferin protein (PRL) / DNA replication licensing factor Mcm7 (MCM7) identical...
262257_at	0.744			AT1G53860 [MIPS, Tair, Tigr]remorin family protein contains Pfam domain, PF03763: Remorin, C-terminal region...
! 257306_at	0.731			AT1G40087 [MIPS, Tair, Tigr][AT1G40087, hypothetical protein]
				AT3G30200 [MIPS, Tair, Tigr][AT3G30200, fructose-6-phosphate 2-kinase-related / fructose-2,6-bisphosphatas...
255714_at	0.820			AT4G00300 [MIPS, Tair, Tigr]fringe-related protein + weak similarity to Fringe [Schistocerca gregaria](Gl:65...
247318_at	0.817			AT5G63990 [MIPS, Tair, Tigr]3'(2'),5'-bisphosphate nucleotidase, putative / inositol polyphosphate 1-phospha...
258521_at	0.863			AT3G06680 [MIPS, Tair, Tigr]60S ribosomal protein L29 (RPL29B) similar to 60S ribosomal protein L29 GB:P2588...
246891_at	0.864			AT5G25490 [MIPS, Tair, Tigr]zinc finger (Ran-binding) family protein contains Pfam domain, PF00641: Zn-finge...
260434_at	0.866			AT1G68330 [MIPS, Tair, Tigr]expressed protein]
262793_at	0.519			AT1G13110 [MIPS, Tair, Tigr]cytochrome P450 71B7 (CYP71B7) identical to (SP:Q96514) cytochrome P450 71B7 [Ar...
245069_at	0.555			AT2G23250 [MIPS, Tair, Tigr]UDP-glucuronosyl/UDP-glucosyl transferase family protein contains similarity to ...
255962_at	0.594			AT1G22335 [MIPS, Tair, Tigr]expressed protein ; expression supported by MPSS]
267017_at	0.977			AT2G39140 [MIPS, Tair, Tigr]pseudouridine synthase family protein]
254795_at	0.960			AT4G12990 [MIPS, Tair, Tigr]expressed protein]
264329_at	0.941			AT1G04140 [MIPS, Tair, Tigr]transducin family protein / WD-40 repeat family protein contains 4 WD-40 repeats...
250562_at	0.925			AT5G08040 [MIPS, Tair, Tigr]expressed protein]
259440_at	1.003			AT1G01550 [MIPS, Tair, Tigr]expressed protein]
257696_at	1.003			AT3G12690 [MIPS, Tair, Tigr]protein kinase, putative similar to viroid symptom modulation protein [Lycopersi...
251957_at	1.015			AT3G63690 [MIPS, Tair, Tigr]zinc finger (C3HC4-type RING finger) family protein contains a Zinc finger, C3HC...
262500_at	1.093			AT1G21760 [MIPS, Tair, Tigr]F-box family protein Contains PF 00646 F-box domain. ESTs gb Z37267, gb R90412, ...
256515_at	1.100			AT1G66020 [MIPS, Tair, Tigr]terpene synthase/cyclase family protein contains Pfam profile: PF01397: Terpene ...
262961_at	1.053			AT1G54490 [MIPS, Tair, Tigr]5'-3' exoribonuclease (XRN4) identical to XRN4 [Arabidopsis thaliana] Gl:1187562...
245314_at	1.227			AT4G16745 [MIPS, Tair, Tigr]exostosin family protein contains Pfam PF03016: Exostosin family]
263217_at	0.233			AT1G30740 [MIPS, Tair, Tigr]FAD-binding domain-containing protein similar to SP P30986 reticuline oxidase pr...
! 260707_s_at	1.691			AT1G32410 [MIPS, Tair, Tigr]vacuolar protein sorting 55 family protein / VPS55 family protein contains Pfam ...
247336_at	1.711			AT5G63640 [MIPS, Tair, Tigr]VHS domain-containing protein / GAT domain-containing protein weak similarity to...
257371_at	1.874			AT2G47810 [MIPS, Tair, Tigr]histone-like transcription factor (CBF/NF-Y) family protein contains Pfam PF0080...
267499_at	1.531			AT2G45500 [MIPS, Tair, Tigr]AAA-type ATPase family protein similar to SP Q9QYY8 Spastin (Fragment) [Mus musc...
260795_at	1.544			AT1G06225 [MIPS, Tair, Tigr]Clavata3 / ESR-Related-3 (CLE3) CLAVATA3/ESR-Related 3 (CLE3)]
266418_at	1.561			AT2G38750 [MIPS, Tair, Tigr]annexin 4 (ANN4) nearly identical to annexin (AnnAt4) [Arabidopsis thaliana] Gl:...
! 267234_x_at	1.495			AT2G43930 [MIPS, Tair, Tigr]protein kinase family protein contains similarity to NPK1-related protein kinase...
262330_at	1.473			AT1G64070 [MIPS, Tair, Tigr]disease resistance protein (TIR-NBS-LRR class), putative domain signature TIR-NB...
256702_at	2.061			AT3G30380 [MIPS, Tair, Tigr]expressed protein ; expression supported by MPSS]
		Hormone: ABA (-)		Control

10 Gene ABA-induziert  
15 Gene ABA-reprimiert  
10 Gene nicht reguliert

## ABA\_Random4:

Treatment			
Hormone: ABA (+)			
Probeset		AGI/Links/Annotation	
250707_at	1.060	AT5G05950	[MIPS, Tair, Tigr]expressed protein]
249372_at	1.048	AT5G40760	[MIPS, Tair, Tigr]glucose-6-phosphate 1-dehydrogenase / G6PD (ACG12) identical to glucose-6-phospha...
251353_at	1.080	AT3G61080	[MIPS, Tair, Tigr]fructosamine kinase family protein contains Pfam PF03881: Fructosamine kinase]
246087_at	1.012	AT5G20580	[MIPS, Tair, Tigr]expressed protein predicted protein, Arabidopsis thaliana]
267303_at	1.121	AT2G30090	[MIPS, Tair, Tigr]GCN5-related N-acetyltransferase (GNAT) family protein contains Pfam profile PF0...
249914_at	1.121	AT5G22850	[MIPS, Tair, Tigr]aspartyl protease family protein contains Pfam profile: PF00026 eukaryotic aspar...
253142_at	1.136	AT4G35520	[MIPS, Tair, Tigr]DNA mismatch repair family protein similar to SP Q9UHC1 DNA mismatch repair prot...
264531_at	1.174	AT1G30790	[MIPS, Tair, Tigr]F-box family protein contains Pfam PF00646: F-box domain; contains TIGRFAM TIGR0...
261477_at	1.168	AT1G14310	[MIPS, Tair, Tigr]haloacid dehalogenase-like hydrolase family protein contains InterPro accession ...
259734_at	1.286	AT1G77500	[MIPS, Tair, Tigr]expressed protein contains Pfam domains, PF04782: Protein of unknown function (D...
267492_at	0.813	AT2G30620	[MIPS, Tair, Tigr]histone H1.2 nearly identical to SP P26569 Histone H1.2 {Arabidopsis thaliana}]
250043_at	0.810	AT5G18430	[MIPS, Tair, Tigr]GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL1 GI:1...
254736_at	0.789	AT4G13820	[MIPS, Tair, Tigr]disease resistance family protein / LRR family protein contains leucine rich-rep...
261043_at	0.833	AT1G01220	[MIPS, Tair, Tigr]GHMP kinase-related contains similarity to L-fucose kinase [Homo sapiens] gi 212...
249589_at	0.833	AT5G37850	[MIPS, Tair, Tigr]pikB-type carbohydrate kinase family protein contains Pfam profile: PF00294 pikB...
266133_at	0.843	AT2G45120	[MIPS, Tair, Tigr]zinc finger (C2H2 type) family protein contains Pfam domain, PF00096: Zinc finge...
250390_at	0.762	AT5G11340	[MIPS, Tair, Tigr]GCN5-related N-acetyltransferase (GNAT) family protein low similarity to SP Q035...
260993_at	0.899	AT1G12140	[MIPS, Tair, Tigr]flavin-containing monooxygenase family protein / FMO family protein similar to f...
255838_at	0.903	AT2G33490	[MIPS, Tair, Tigr]hydroxyproline-rich glycoprotein family protein Common family member:At3g26910 [...
246316_at	0.917	AT3G56890	[MIPS, Tair, Tigr]F-box family protein-related predicted proteins - Arabidopsis thaliana; contains...
249184_at	0.880	AT5G43020	[MIPS, Tair, Tigr]leucine-rich repeat transmembrane protein kinase, putative]
262616_at	0.936	AT1G06620	[MIPS, Tair, Tigr]2-oxoglutarate-dependent dioxygenase, putative similar to 2A6 (GI:599622) and to...
246006_at	0.936	AT5G08290	[MIPS, Tair, Tigr]yellow-leaf-specific protein 8 (YLS8) / mitosis protein DIM1, putative contains ...
! 255908_s	0.947	AT1G18000	[MIPS, Tair, Tigr][[AT1G18000, expressed protein contains 11 transmembrane domains;]]
267433_at	0.959	AT1G18010	[MIPS, Tair, Tigr][AT1G18010, expressed protein contains 11 transmembrane domains;]
255244_at	0.621	AT2G26350	[MIPS, Tair, Tigr]zinc-binding peroxisomal integral membrane protein (PEX10) identical to zinc-bin...
249188_at	0.610	AT4G05620	[MIPS, Tair, Tigr]F-box family protein similar to SKP1 interacting partner 6 [Arabidopsis thaliana...
262362_at	0.572	AT5G42830	[MIPS, Tair, Tigr]transferase family protein similar to anthranilate N-hydroxycinnamoyl/benzoyltra...
250805_at	0.663	AT1G72840	[MIPS, Tair, Tigr]disease resistance protein (TIR-NBS-LRR class), putative domain signature TIR-NB...
! 263832_at	0.510	AT5G05050	[MIPS, Tair, Tigr]peptidase C1A papain family protein weak similarity to berghepain-2 [Plasmodium ...
262733_s	0.526	AT2G15450	[MIPS, Tair, Tigr][AT2G15450, glycoside hydrolase family 28 protein / polygalacturonase (pectina...
249032_at	0.487	AT2G15460	[MIPS, Tair, Tigr][[AT2G15460, glycoside hydrolase family 28 protein / polygalacturonase (pectina...
262751_at	1.622	AT2G40310	[MIPS, Tair, Tigr][AT2G40310, glycoside hydrolase family 28 protein / polygalacturonase (pectina...
256577_at	1.775	AT4G13760	[MIPS, Tair, Tigr][AT4G13760, glycoside hydrolase family 28 protein / polygalacturonase (pectina...
261538_at	1.973	AT1G28670	[MIPS, Tair, Tigr]lipase identical to lipase GB:AAA93262 GI:1145627 [Arabidopsis thaliana] (FEBS L...
		AT5G44910	[MIPS, Tair, Tigr]Toll-Interleukin-Resistance (TIR) domain-containing protein domain signature TIR...
		AT1G16310	[MIPS, Tair, Tigr]cation efflux family protein contains cation efflux family protein domain, Pfam:...
		AT3G28220	[MIPS, Tair, Tigr]meprin and TRAF homology domain-containing protein / MATH domain-containing prot...
		AT1G01830	[MIPS, Tair, Tigr]armadillo/beta-catenin repeat family protein armadillo/beta-catenin-like repeats...
Hormone: ABA (-)			

Control

6 Gene ABA-induziert  
14 Gene ABA-reprimiert  
15 Gene nicht reguliert

## ABA\_Random5:

Treatment		Hormone: ABA (+)		AGI/Links/Annotation	
Probeset					
259129_a	2.523	AT3G02150	[MIPS, Tair, Tigr]	TCP family transcription factor, putative similar to transcription factor PCF6 [...]	
257056_a	2.522	AT3G15350	[MIPS, Tair, Tigr]	glycosyltransferase family 14 protein / core-2/I-branching enzyme family protein...	
258686_a	2.418	AT3G07840	[MIPS, Tair, Tigr]	polygalacturonase, putative / pectinase, putative strong similarity to polygalac...	
260228_a	2.932	AT1G74540	[MIPS, Tair, Tigr]	cytochrome P450, putative similar to cytochrome P450 GB:O48922 [Glycine max]	
! 256451_s	3.667	AT1G75170	[MIPS, Tair, Tigr]	SEC14 cytosolic factor family protein / phosphoglyceride transfer family protein...	
259419_a	0.737	AT1G06310	[MIPS, Tair, Tigr]	acyl-CoA oxidase, putative strong similarity to acyl-CoA oxidase ACX3 GI:8163758...	
253780_a	0.737	AT4G28400	[MIPS, Tair, Tigr]	protein phosphatase 2C, putative / PP2C, putative protein phosphatase 2C-fission...	
245984_a	0.756	AT5G13090	[MIPS, Tair, Tigr]	expressed protein predicted proteins - Arabidopsis thaliana]	
260258_a	0.791	AT1G74270	[MIPS, Tair, Tigr]	60S ribosomal protein L35a (RPL35aC) similar to ribosomal protein L33B GB:NP_014...	
262663_a	0.850	AT1G13940	[MIPS, Tair, Tigr]	expressed protein]	
250111_a	0.849	AT5G16290	[MIPS, Tair, Tigr]	acetolactate synthase small subunit, putative similar to gi:5931761 from Nicotia...	
251267_a	0.852	AT3G62330	[MIPS, Tair, Tigr]	zinc knuckle (CCHC-type) family protein contains Pfam domain, PF00098: Zinc knuc...	
245599_a	0.842	AT4G14220	[MIPS, Tair, Tigr]	zinc finger (C3HC4-type RING finger) family protein contains Pfam profile: PF000...	
248554_a	0.868	AT5G50330	[MIPS, Tair, Tigr]	ABC1 family protein contains Pfam domain, PF03109: ABC1 family]	
263459_a	0.826	AT2G31820	[MIPS, Tair, Tigr]	ankyrin repeat family protein contains ankyrin repeat domains, Pfam:PF00023]	
247581_a	0.891	AT5G61350	[MIPS, Tair, Tigr]	protein kinase family protein contains protein kinase domain, Pfam:PF00069]	
260962_a	0.975	AT1G44980	[MIPS, Tair, Tigr]	pectinesterase family protein contains Pfam profile: PF01095 pectinesterase]	
254484_a	0.971	AT4G20760	[MIPS, Tair, Tigr]	short-chain dehydrogenase/reductase (SDR) family protein contains Pfam profile: ...	
262467_a	0.963	AT1G50240	[MIPS, Tair, Tigr]	armadillo/beta-catenin repeat family protein contains Pfam profile: PF00514 arma...	
253108_a	0.988	AT4G35900	[MIPS, Tair, Tigr]	DNA-binding protein-related weak similarity to DNA-binding factor gmlip15 [Zea m...	
! 250667_a	0.983	AT5G07090	[MIPS, Tair, Tigr]	[AT5G07090, 40S ribosomal protein S4 (RPS4B)]	
		AT5G58420	[MIPS, Tair, Tigr]	[AT5G58420, 40S ribosomal protein S4 (RPS4D) ribosomal protein S4, Arabidopsis...	
267347_a	0.994	AT2G39950	[MIPS, Tair, Tigr]	expressed protein]	
265659_a	1.006	AT2G25440	[MIPS, Tair, Tigr]	leucine-rich repeat family protein contains leucine rich-repeat (LRR) domains Pf...	
248492_a	1.019	AT5G51040	[MIPS, Tair, Tigr]	expressed protein]	
264366_a	1.076	AT1G03250	[MIPS, Tair, Tigr]	expressed protein]	
252210_a	1.085	AT3G50410	[MIPS, Tair, Tigr]	Dof-type zinc finger domain-containing protein]	
258591_a	0.667	AT3G04360	[MIPS, Tair, Tigr]	C2 domain-containing protein low similarity to cold-regulated gene SRC2 [Glycine...	
256362_a	0.625	AT1G66450	[MIPS, Tair, Tigr]	DC1 domain-containing protein contains Pfam profile PF03107: DC1 domain]	
! 259204_s	0.564	AT1G34610	[MIPS, Tair, Tigr]	[AT1G34610, Ulp1 protease family protein contains Pfam profile PF02902: Ulp1 p...	
		AT3G09170	[MIPS, Tair, Tigr]	[AT3G09170, Ulp1 protease family protein contains Pfam profile PF02902: Ulp1 pr...	
246919_a	0.581	AT5G25460	[MIPS, Tair, Tigr]	expressed protein contains Pfam profile PF04862: Protein of unknown function, DU...	
263174_a	1.239	AT1G54040	[MIPS, Tair, Tigr]	kelch repeat-containing protein contains Pfam PF01344: Kelch motif (4 repeats); ...	
250803_a	1.234	AT5G04980	[MIPS, Tair, Tigr]	endonuclease/exonuclease/phosphatase family protein contains similarity to inosi...	
249578_a	1.264	AT5G37730	[MIPS, Tair, Tigr]	expressed protein]	
259473_a	1.314	AT1G19025	[MIPS, Tair, Tigr]	DNA cross-link repair protein-related contains weak similarity to Swiss-Prot:P30...	
264836_a	1.529	AT1G03610	[MIPS, Tair, Tigr]	expressed protein]	
		Hormone: ABA (-)			
Control					

10 Gene ABA-induziert  
14 Gene ABA-reprimiert  
11 Gene nicht reguliert

### 7.3 ABF1-Downstreamgruppe und Vergleichsgruppen zufällig ausgewählter Gene

ABF1-Downstreamgruppe:

Treatment	
Hormone: ABA (+)	
Probeset	AGI/Links/Annotation
253604_at1.101	AT4G30970 [MIPS, Tair, Tigr]hypothetical protein]
252951_at1.043	AT4G38700 [MIPS, Tair, Tigr]disease resistance-responsive family protein related to disease resistance respo...
261752_at0.955	AT1G76290 [MIPS, Tair, Tigr]AMP-dependent synthetase and ligase family protein similar to AMP-binding protei...
261518_at0.943	AT1G71695 [MIPS, Tair, Tigr]peroxidase 12 (PER12) (P12) (PRXR6) identical to SP Q96520 Peroxidase 12 precurs...
251785_at0.980	AT3G55130 [MIPS, Tair, Tigr]ABC transporter family protein breast cancer resistance protein 1 BCRP1, Mus mus...
265620_at0.907	AT2G27310 [MIPS, Tair, Tigr]F-box family protein contains Pfam PF00646: F-box domain;; similar to SKP1 inter...
254186_at1.247	AT4G24010 [MIPS, Tair, Tigr]cellulose synthase family protein similar to Zea mays cellulose synthase-5 [gi:9...
261772_at0.680	AT1G76240 [MIPS, Tair, Tigr]expressed protein contains Pfam profile PF03087: Arabidopsis protein of unknown ...
250010_at0.677	AT5G18450 [MIPS, Tair, Tigr]AP2 domain-containing transcription factor, putative DREB2A, Arabidopsis thalian...
!! 251112_s_at0.640	AT5G01320 [MIPS, Tair, Tigr][AT5G01320, pyruvate decarboxylase, putative strong similarity to pyruvate deca...
261913_at0.790	AT5G01330 [MIPS, Tair, Tigr][AT5G01330, pyruvate decarboxylase, putative strong similarity to pyruvate dec...
257467_at0.466	AT1G65860 [MIPS, Tair, Tigr]flavin-containing monooxygenase family protein / FMO family protein similar to f...
257974_at1.597	AT1G31320 [MIPS, Tair, Tigr]LOB domain protein 4 / lateral organ boundaries domain protein 4 (LBD4) identica...
253358_at1.714	AT3G20820 [MIPS, Tair, Tigr]leucine-rich repeat family protein contains similarity to Cf-2.1 [Lycopersicon p...
263176_at1.916	AT4G32940 [MIPS, Tair, Tigr]vacuolar processing enzyme gamma / gamma-VPE nearly identical to SP Q39119 Vacuo...
265119_at13.932	AT1G05530 [MIPS, Tair, Tigr]UDP-glucuronosyl/UDP-glucosyl transferase family protein contains Pfam profile: ...
	AT1G62570 [MIPS, Tair, Tigr]flavin-containing monooxygenase family protein / FMO family protein low similari...
Hormone: ABA (-)	
Control	

5 Gene ABA-induziert  
5 Gene ABA-reprimiert  
6 Gene nicht reguliert

At2g06908 nicht auf dem Chip  
At1g63040 nicht auf dem Chip  
At3g10510 nicht auf dem Chip  
At5g04880 nicht auf dem Chip  
At5g16370 nicht auf dem Chip



## Vergleichsgruppen für ABF1-Downstreamgruppe

Down1:

Treatment		
Hormone: ABA (+)		
Probeset	AGI/Links/Annotation	
258746_at1.032	AT3G05950 [MIPS, Tair, Tigr]germin-like protein, putative similar to germin-like protein GLP6 [SP P92997]	
247372_at1.051	AT5G63280 [MIPS, Tair, Tigr]zinc finger (C2H2 type) family protein contains Pfam domain, PF00096: Zinc finge...	
!! 257590_s_at0.972	AT3G24870 [MIPS, Tair, Tigr][AT3G24870, expressed protein]	
	AT3G24880 [MIPS, Tair, Tigr][AT3G24880, expressed protein]	
251018_at0.967	AT5G02450 [MIPS, Tair, Tigr]60S ribosomal protein L36 (RPL36C) 60S ribosomal protein L36, Arabidopsis thalia...	
246543_at1.145	AT5G15060 [MIPS, Tair, Tigr]hypothetical protein]	
256079_at0.819	AT1G20680 [MIPS, Tair, Tigr]expressed protein contains Pfam profile PF05742: Protein of unknown function (DU...	
245950_at0.818	AT5G19570 [MIPS, Tair, Tigr]expressed protein]	
265027_at0.813	AT1G24450 [MIPS, Tair, Tigr]ribonuclease III family protein contains similarity to Swiss-Prot:P51837 ribonuc...	
258531_at0.804	AT3G06720 [MIPS, Tair, Tigr]importin alpha-1 subunit, putative (IMPA1) similar to importin alpha-1 subunit (...	
!! 262717_s_at0.788	AT1G16400 [MIPS, Tair, Tigr][AT1G16400, cytochrome P450 family protein similar to gb AF069494 cytochrome P...	
	AT1G16410 [MIPS, Tair, Tigr][AT1G16410, cytochrome P450, putative similar to gb AF069494 cytochrome P450 fr...	
260607_at0.903	AT2G43700 [MIPS, Tair, Tigr]lectin protein kinase family protein contains pfam domains PF00139: Legume lecti...	
254930_at0.887	AT4G11450 [MIPS, Tair, Tigr]expressed protein]	
263632_at0.693	AT2G04795 [MIPS, Tair, Tigr]expressed protein]	
251748_at0.734	AT3G55680 [MIPS, Tair, Tigr]invertase/pectin methylesterase inhibitor family protein contains Pfam profile P...	
266921_at1.294	AT2G45970 [MIPS, Tair, Tigr]cytochrome P450, putative]	
256261_at1.520	AT3G12160 [MIPS, Tair, Tigr]Ras-related GTP-binding family protein similar to ras-related GTP-binding protei...	
Hormone: ABA (-)		
		Control

2 ABA-induzierte Gene  
 7 ABA-reprimierte Gene  
 7 nicht regulierte Gene

Down2:

Treatment		
Hormone: ABA (+)		
Probeset		AGI/Links/Annotation
263420_a	0.680	AT2G17240 [MIPS, Tair, Tigr]expressed protein]
261500_a	0.683	AT1G28400 [MIPS, Tair, Tigr]expressed protein similar to E6 (GI:1000090) [Gossypium barbadense]]
262039_a	0.646	AT1G80050 [MIPS, Tair, Tigr]adenine phosphoribosyltransferase 2 (APT2) identical to SP Q42563 Adenine phosph...
262855_a	0.733	AT1G20860 [MIPS, Tair, Tigr]phosphate transporter family protein similar to phosphate transporter [Catharant...
249378_a	0.750	AT5G40450 [MIPS, Tair, Tigr]expressed protein]
245670_a	0.770	AT1G28210 [MIPS, Tair, Tigr]DNAJ heat shock protein, putative strong similarity to mitochondrial DnaJ protei...
265683_a	0.815	AT2G24400 [MIPS, Tair, Tigr]auxin-responsive protein, putative / small auxin up RNA (SAUR_D) similar to SAUR...
248634_a	0.842	AT5G49030 [MIPS, Tair, Tigr]tRNA synthetase class I (I, L, M and V) family protein similar to SP P41972 Isol...
255669_a	0.879	AT4G00416 [MIPS, Tair, Tigr]methyl-CpG-binding domain-containing protein contains Pfam profile PF01429: Meth...
247973_a	0.887	AT5G56770 [MIPS, Tair, Tigr]hypothetical protein]
257497_a	0.867	AT1G51430 [MIPS, Tair, Tigr]expressed protein]
265638_a	1.022	AT2G27340 [MIPS, Tair, Tigr]hypothetical protein]
260806_a	0.963	AT1G78260 [MIPS, Tair, Tigr]RNA recognition motif (RRM)-containing protein similar to RNA recognition motif-...
247155_a	1.105	AT5G65750 [MIPS, Tair, Tigr]2-oxoglutarate dehydrogenase E1 component, putative / oxoglutarate decarboxylase...
256546_a	1.373	AT3G14820 [MIPS, Tair, Tigr]GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL3 GI:1...
251412_a	2.372	AT3G60220 [MIPS, Tair, Tigr]zinc finger (C3HC4-type RING finger) family protein (ATL4) contains Pfam profile...
Hormone: ABA (-)		
Control		

2 ABA-induzierte Gene  
9 ABA-reprimierte Gene  
5 nicht regulierte Gene

Treatment		
Probeset	Hormone: ABA (+)	AGI/Links/Annotation
	257579_a	1.477
253165_a	1.539	AT4G35320 [MIPS, Tair, Tigr]expressed protein predicted protein, Arabidopsis thaliana]
266185_a	1.282	AT2G38950 [MIPS, Tair, Tigr]transcription factor jumonji (jmi) family protein / zinc finger (C5HC2 type) fam...
261602_a	0.951	AT1G49630 [MIPS, Tair, Tigr]peptidase M16 family protein / insulinase family protein contains Pfam domain, P...
248569_a	0.976	AT5G49770 [MIPS, Tair, Tigr]leucine-rich repeat transmembrane protein kinase, putative]
255583_a	0.904	AT4G01510 [MIPS, Tair, Tigr]transmembrane ARV1-like family protein contains similarity to ARV1 [Homo sapiens...
267144_a	1.014	AT2G38110 [MIPS, Tair, Tigr]phospholipid/glycerol acyltransferase family protein low similarity to SP O87707...
260064_a	1.037	AT1G73730 [MIPS, Tair, Tigr]ethylene-insensitive3-like3 (EIL3) identical to ethylene-insensitive3-like3 (EIL...
266958_a	1.113	AT2G34630 [MIPS, Tair, Tigr]geranyl diphosphate synthase, putative / GPPS, putative / dimethylallyltransfera...
266980_a	0.835	AT2G39390 [MIPS, Tair, Tigr]60S ribosomal protein L35 (RPL35B)]
262663_a	0.850	AT1G13940 [MIPS, Tair, Tigr]expressed protein]
267285_a	0.809	AT2G23690 [MIPS, Tair, Tigr]expressed protein]
245044_a	0.810	AT2G26500 [MIPS, Tair, Tigr]cytochrome b6f complex subunit (petM), putative nearly identical to cytochrome b...
257877_a	0.699	AT3G17140 [MIPS, Tair, Tigr]invertase inhibitor-related weak similarity to tomato invertase inhibitor [Lycop...
253842_a	0.719	AT4G27860 [MIPS, Tair, Tigr]integral membrane family protein contains Pfam PF01988: Integral membrane protei...
253053_a	2.299	AT4G37470 [MIPS, Tair, Tigr]hydrolase, alpha/beta fold family protein low similarity to SP Q59093 3-oxoadipa...
	Hormone: ABA (-)	

139

Down4:

Treatment		
Hormone: ABA (+)		
Probeset	AGI/Links/Annotation	
258119_a	1.561	AT3G14720 [MIPS, Tair, Tigr]mitogen-activated protein kinase, putative / MAPK, putative (MPK19) identical to...
253798_a	1.265	AT4G28500 [MIPS, Tair, Tigr]no apical meristem (NAM) family protein contains Pfam PF02365: No apical meriste...
252180_a	0.966	AT3G50630 [MIPS, Tair, Tigr]kip-related protein 2 (KRP2) / cyclin-dependent kinase inhibitor 2 (ICK2) / cdc2...
250441_a	0.964	AT5G10540 [MIPS, Tair, Tigr]peptidase M3 family protein / thimet oligopeptidase family protein similar to SP...
260438_a	0.972	AT1G68290 [MIPS, Tair, Tigr]bifunctional nuclease, putative similar to bifunctional nuclease [Zinnia elegans...
264896_a	0.988	AT1G23210 [MIPS, Tair, Tigr]glycosyl hydrolase family 9 protein similar to endo-1,4-beta-glucanase GB:CAA671...
262206_a	1.042	AT2G01090 [MIPS, Tair, Tigr]ubiquinol-cytochrome C reductase complex 7.8 kDa protein, putative / mitochondri...
258047_a	0.881	AT3G21240 [MIPS, Tair, Tigr]4-coumarate--CoA ligase 2 / 4-coumaroyl-CoA synthase 2 (4CL2) identical to SP Q9...
246539_a	0.868	AT5G15460 [MIPS, Tair, Tigr]expressed protein]
257913_a	0.811	AT3G25540 [MIPS, Tair, Tigr]longevity-assurance (LAG1) family protein similar to Alternaria stem canker resi...
245526_a	0.818	AT4G15950 [MIPS, Tair, Tigr]expressed protein]
249900_a	0.685	AT5G22640 [MIPS, Tair, Tigr]MORN (Membrane Occupation and Recognition Nexus) repeat-containing protein conta...
246770_a	0.687	AT5G27460 [MIPS, Tair, Tigr]pentatricopeptide (PPR) repeat-containing protein low similarity to DNA-binding ...
254043_a	0.669	AT4G25990 [MIPS, Tair, Tigr]expressed protein]
260331_a	0.705	AT1G80270 [MIPS, Tair, Tigr]DNA-binding protein, putative similar to DNA-binding protein [Triticum aestivum]...
262620_a	0.740	AT1G06540 [MIPS, Tair, Tigr]expressed protein]
Hormone: ABA (-)		
Control		

2 ABA-induzierte Gene  
8 ABA-reprimierte Gene  
6 nicht regulierte Gene

Down5:

Treatment	
Hormone: ABA (+)	
Probeset	AGI/Links/Annotation
265698_at1.412	AT2G32160 [MIPS, Tair, Tigr]expressed protein]
255990_at1.631	AT1G29870 [MIPS, Tair, Tigr]tRNA synthetase class II (G, H, P and S) family protein similar to SP O23627 Gly...
262200_at0.993	AT2G01070 [MIPS, Tair, Tigr]expressed protein similar to membrane protein PTM1 precursor isolog GB:AAB65479]
258646_at0.977	AT3G08040 [MIPS, Tair, Tigr]MATE efflux family protein low similarity to enhanced disease susceptibility 5 [...
250685_at1.020	AT5G06670 [MIPS, Tair, Tigr]kinesin motor protein-related]
245525_at1.022	AT4G15930 [MIPS, Tair, Tigr]dynein light chain, putative similar to dynein light chain 2 [Mus musculus] Gl:1...
245815_at0.938	AT1G26090 [MIPS, Tair, Tigr]hypothetical protein]
!! 259873_s_at1.096	AT1G76720 [MIPS, Tair, Tigr][AT1G76720, eukaryotic translation initiation factor 2 family protein / eIF-2 ...
	AT1G76810 [MIPS, Tair, Tigr][AT1G76810, eukaryotic translation initiation factor 2 family protein / eIF-2 ...
	AT1G76825 [MIPS, Tair, Tigr][AT1G76825, eukaryotic translation initiation factor 2 family protein / eIF-2 f...
245279_at1.103	AT4G17270 [MIPS, Tair, Tigr]Mo25 family protein similar to MO25 protein (early mouse development protein fam...
265725_at1.068	AT2G32030 [MIPS, Tair, Tigr]GCN5-related N-acetyltransferase (GNAT) family protein contains Pfam profile PF0...
250093_at1.147	AT5G17370 [MIPS, Tair, Tigr]WD-40 repeat family protein contains 1 significant, 2 weak WD-40 repeats (PF0040...
261456_at0.569	AT1G21050 [MIPS, Tair, Tigr]expressed protein contains Pfam profile PF04759: Protein of unknown function, DU...
252054_at0.520	AT3G52540 [MIPS, Tair, Tigr]ovate family protein 52% similar to ovate protein (Gl:23429649) [Lycopersicon es...
247540_at0.175	AT5G61590 [MIPS, Tair, Tigr]AP2 domain-containing transcription factor family protein contains Pfam PF00847:...
256933_at2.357	AT3G22600 [MIPS, Tair, Tigr]protease inhibitor/seed storage/lipid transfer protein (LTP) family protein cont...
246926_at1.994	AT5G25240 [MIPS, Tair, Tigr]expressed protein]
Hormone: ABA (-)	
Control	

4 ABA-induzierte Gene  
3 ABA-reprimierte Gene  
9 nicht regulierte Gene

## 7.4 Gene der ABF1\_O2\_5\_50bp Prädiktionsgruppe (ohne redundante Gene, die in mehreren ABF1-O2-Gruppen auftreten)

Treatment		AGI/Links/Annotation	
Probeset	Hormone: ABA (+)		
257953_at1.845		AT3G21865	[MIPS, Tair, Tigr]expressed protein]
249039_at1.787		AT5G44310	[MIPS, Tair, Tigr]late embryogenesis abundant domain-containing protein / LEA domain-containing pr...
265325_at2.022		AT2G18240	[MIPS, Tair, Tigr]RER1 protein, putative similar to SP O48671 RER1B protein (AtRER1B) (Arabidopsis...
250141_at1.978		AT5G14640	[MIPS, Tair, Tigr]protein kinase family protein similar to glycogen synthase kinase-3 homolog Msk-...
! 258501_at2.414		AT3G06780	[MIPS, Tair, Tigr][AT3G06780, glycine-rich protein]
		AT3G06790	[MIPS, Tair, Tigr][AT3G06790, plastid developmental protein DAG, putative similar to DAG protein...
! 248336_at2.348		AT5G52420	[MIPS, Tair, Tigr][AT5G52420, expressed protein]]
		AT5G52430	[MIPS, Tair, Tigr][AT5G52430, hydroxyproline-rich glycoprotein family protein Common family membe...
260797_at1.671		AT1G78390	[MIPS, Tair, Tigr]9-cis-epoxycarotenoid dioxygenase, putative / neoxanthin cleavage enzyme, putati...
257087_at1.652		AT3G20500	[MIPS, Tair, Tigr]calciueurin-like phosphoesterase family protein contains Pfam profile: PF00149 c...
246126_at1.621		AT5G20070	[MIPS, Tair, Tigr]MutT/nudix family protein low similarity to SPIQ19427 NADH pyrophosphatase (EC 3...
261269_at1.525		AT1G26690	[MIPS, Tair, Tigr]emp24/gp25L/p24 family protein similar to SPIP49755 Transmembrane protein Tmp21 ...
259965_at1.369		AT1G53670	[MIPS, Tair, Tigr]transcription factor-related similar to pilin-like transcription factor [Homo sa...
255011_at1.370		AT4G10040	[MIPS, Tair, Tigr]cytochrome c, putative similar to cytochrome c [Pumpkin, Winter squash] SWISS-PR...
251708_at1.372		AT3G56580	[MIPS, Tair, Tigr]zinc finger (C3HC4-type RING finger) family protein contains INTERPRO domain, IP...
253322_at1.359		AT4G33980	[MIPS, Tair, Tigr]expressed protein]
248665_at1.441		AT5G48655	[MIPS, Tair, Tigr]zinc finger (C3HC4-type RING finger) family protein contains Pfam profile: PF000...
265630_at1.262		AT2G27350	[MIPS, Tair, Tigr]OTU-like cysteine protease family protein contains Pfam profile PF02338: OTU-lik...
264676_at1.271		AT1G09870	[MIPS, Tair, Tigr]histidine acid phosphatase family protein contains Pfam profile PF00328: Histi...
250601_at0.952		AT5G07810	[MIPS, Tair, Tigr]SNF2 domain-containing protein / helicase domain-containing protein / HNH endonu...
! 246627_s_at0.937		AT1G48860	[MIPS, Tair, Tigr][AT1G48860, 3-phosphoshikimate 1-carboxyvinyltransferase, putative / 5-enolpyru...
		AT2G45300	[MIPS, Tair, Tigr][AT2G45300, 3-phosphoshikimate 1-carboxyvinyltransferase / 5-enolpyruvylshikim...
254658_at0.986		AT4G18230	[MIPS, Tair, Tigr]expressed protein predicted gene M02B7.4, Caenorhabditis elegans, PID:G1572781]
245242_at0.994		AT1G44446	[MIPS, Tair, Tigr]chlorophyll a oxygenase (CAO) / chlorophyll b synthase identical to chlorophyll ...
265415_at1.033		AT2G20890	[MIPS, Tair, Tigr]expressed protein]
261904_at1.035		AT1G65040	[MIPS, Tair, Tigr]zinc finger (C3HC4-type RING finger) family protein contains Pfam profile: PF000...
255300_at1.048		AT4G04870	[MIPS, Tair, Tigr]CDP-alcohol phosphatidyltransferase family protein similar to SPIQ07560 Cardioli...
263769_at0.857		AT2G06390	[MIPS, Tair, Tigr]hypothetical protein]
256224_at0.868		AT1G56330	[MIPS, Tair, Tigr]GTP-binding protein (SAR1B) identical to GTP-binding protein (SAR1B) [Arabidopsi...
266332_at1.107		AT2G32415	[MIPS, Tair, Tigr]3'-5' exonuclease domain-containing protein similar to SPIQ12149 Exosome complex...
245289_at1.113		AT4G15470	[MIPS, Tair, Tigr]expressed protein low similarity to N-methyl-D-aspartate receptor-associated pro...
259476_at1.139		AT1G19000	[MIPS, Tair, Tigr]myb family transcription factor similar to MybSt1 GI:7705206 from [Solanum tuber...
256680_at1.182		AT3G52230	[MIPS, Tair, Tigr]expressed protein]
250400_at1.162		AT5G10740	[MIPS, Tair, Tigr]protein phosphatase 2C-related / PP2C-related protein phosphatase 2C, alfalfa, P...
260144_at3.376		AT1G71960	[MIPS, Tair, Tigr]ABC transporter family protein similar to breast cancer resistance protein GB:AA...
247488_at5.806		AT5G61820	[MIPS, Tair, Tigr]expressed protein MtN19, Medicago truncatula, EMBL:MTY15367]
252321_at10.557		AT3G48510	[MIPS, Tair, Tigr]expressed protein]
247437_at33.736		AT5G62490	[MIPS, Tair, Tigr]ABA-responsive protein (HVA22b) identical to AtHVA22b [Arabidopsis thaliana] GI:...
	Hormone: ABA (-)		
			Control

23 ABA-induzierte Gene  
2 ABA-reprimierte Gene  
10 nicht regulierte Gene

## 7.5 Gene der ABF1\_O2\_51\_100bp Prädiktionsgruppe (ohne redundante Gene, die in mehreren ABF1-O2-Gruppen auftreten)

Treatment	
Probeset	Hormone: ABA (+)
AGI/Links/Annotation	
264019_at.2.179	AT2G21130 [MIPS, Tair, Tigr]peptidyl-prolyl cis-trans isomerase / cyclophilin (CYP2) / rotamase identical to...
245483_at.2.239	AT4G16190 [MIPS, Tair, Tigr]cysteine proteinase, putative contains similarity to papain-like cysteine protei...
258575_at.2.489	AT3G04240 [MIPS, Tair, Tigr]O-linked N-acetyl glucosamine transferase, putative similar to O-GlcNAc transfer...
245602_at.2.486	AT4G14270 [MIPS, Tair, Tigr]expressed protein]
253590_at.2.970	AT4G30850 [MIPS, Tair, Tigr]expressed protein contains Pfam domain, PF03006: Uncharacterised protein family ...
!! 254050_s_at.3.623	AT4G25670 [MIPS, Tair, Tigr][[AT4G25670, expressed protein]]
	AT4G25690 [MIPS, Tair, Tigr][AT4G25690, expressed protein]
252089_at.1.116	AT3G52110 [MIPS, Tair, Tigr]expressed protein]
245096_at.1.106	AT2G40880 [MIPS, Tair, Tigr]cysteine protease inhibitor, putative / cystatin, putative (FL3-27) similar to P...
261938_at.1.157	AT1G22510 [MIPS, Tair, Tigr]zinc finger (C3HC4-type RING finger) family protein contains Pfam profile: PF000...
253221_at.0.965	AT4G34870 [MIPS, Tair, Tigr]peptidyl-prolyl cis-trans isomerase / cyclophilin (CYP1) / rotamase identical to...
252130_at.0.958	AT3G50820 [MIPS, Tair, Tigr]oxygen-evolving enhancer protein, chloroplast, putative / 33 kDa subunit of oxyg...
265073_at.0.988	AT1G55480 [MIPS, Tair, Tigr]expressed protein]
251147_at.0.985	AT3G63480 [MIPS, Tair, Tigr]kinesin heavy chain, putative kinesin heavy chain, Syncephalastrum racemosum, SW...
262166_at.1.018	AT1G74840 [MIPS, Tair, Tigr]myb family transcription factor similar to myb-related transcription activator G...
249111_at.1.011	AT5G43770 [MIPS, Tair, Tigr]proline-rich family protein contains proline-rich extensin domains, INTERPRO:IPR...
264096_at.0.794	AT1G78995 [MIPS, Tair, Tigr]expressed protein]
245981_at.0.864	AT5G13100 [MIPS, Tair, Tigr]expressed protein]
261783_at.1.247	AT1G08190 [MIPS, Tair, Tigr]vacuolar assembly protein, putative (VPS41) 99.8% identical to Vacuolar assembly...
245099_at.1.333	AT2G40830 [MIPS, Tair, Tigr]zinc finger (C3HC4-type RING finger) family protein contains Pfam domain, PF0009...
263754_at.1.469	AT2G21510 [MIPS, Tair, Tigr]DNAJ heat shock N-terminal domain-containing protein similar to SPIP39101 CAJ1 p...
258376_at.1.484	AT3G17680 [MIPS, Tair, Tigr]expressed protein similar to GB:AAD49756 from [Arabidopsis thaliana]]
266090_at.1.427	AT2G38000 [MIPS, Tair, Tigr]chaperone protein dnaJ-related weak similarity to Chaperone protein dnaJ (Swiss-...
267321_at.1.538	AT2G19320 [MIPS, Tair, Tigr]expressed protein]
262116_at.5.674	AT1G02816 [MIPS, Tair, Tigr]expressed protein contains Pfam profile PF04398: Protein of unknown function, DU...
Probeset	Hormone: ABA (-)
Control	

14 ABA-induzierte Gene  
2 ABA-reprimierte Gene  
8 nicht regulierte Gene

Nicht auf dem Chip:  
At3g12970  
At5g15970

## 7.6 Gene der ABF1\_O2\_101\_150bp Prädiktionsgruppe (ohne redundante Gene, die in mehreren ABF1-O2-Gruppen auftreten)

Treatment			
Hormone: ABA (+)			
Probeset		AGI/Links/Annotation	
255418_at	1.431	AT4G03200	[MIPS, Tair, Tigr]expressed protein contains Pfam PF03190: Protein of unknown function, DUF255]
250465_at	1.457	AT5G10070	[MIPS, Tair, Tigr]RNase L inhibitor protein-related contains Pfam profiles PF04034: Domain of unk...
262499_at	1.239	AT1G21770	[MIPS, Tair, Tigr]expressed protein]
259537_at	1.565	AT1G12370	[MIPS, Tair, Tigr]type II CPD photolyase PHR1 (PHR1) nearly identical to type II CPD photolyase PH...
247944_at	1.645	AT5G57100	[MIPS, Tair, Tigr]transporter-related low similarity to GDP-fucose transporter [Caenorhabditis ele...
258682_at	1.736	AT3G08720	[MIPS, Tair, Tigr]serine/threonine protein kinase (PK19) identical to serine/threonine-protein kin...
265149_at	0.897	AT1G51400	[MIPS, Tair, Tigr]photosystem II 5 kD protein 100% identical to Gl:4836947 (F5D21.10)]
252105_at	0.969	AT3G51470	[MIPS, Tair, Tigr]protein phosphatase 2C, putative / PP2C, putative protein phosphatase-2C, Mesemb...
264092_at	0.807	AT1G79040	[MIPS, Tair, Tigr]photosystem II 10 kDa polypeptide identical to photosystem II 10 kDa polypeptide...
252293_at	0.844	AT3G48990	[MIPS, Tair, Tigr]AMP-dependent synthetase and ligase family protein similar to peroxisomal-coenzy...
252430_at	0.768	AT3G47470	[MIPS, Tair, Tigr]chlorophyll A-B binding protein 4, chloroplast / LHCl type III CAB-4 (CAB4) iden...
260543_at	2.474	AT2G43330	[MIPS, Tair, Tigr]sugar transporter family protein similar to SP Q96QE2 Proton myo-inositol co-tra...
248750_at	7.193	AT5G47530	[MIPS, Tair, Tigr]auxin-responsive protein, putative similar to auxin-induced protein AIR12 (Gl:11...
Hormone: ABA (-)			
		Control	

8 ABA-induzierte Gene  
 3 ABA-reprimierte Gene  
 2 nicht regulierte Gene

Nicht auf dem Chip:  
 At5g38590



Treatment		
Hormone: ABA (+)		
Probeset	AGI/Links/Annotation	
265097_at1.393	AT1G04020 [MIPS, Tair, Tigr]zinc finger (C3HC4-type RING finger) family protein / BRCT domain-containing pro...	
247033_at1.498	AT5G67250 [MIPS, Tair, Tigr]SKP1 interacting partner 2 (SKIP2) identical to SKP1 interacting partner 2 GI:10...	
261151_at1.264	AT1G19650 [MIPS, Tair, Tigr]SEC14 cytosolic factor, putative / phosphoglyceride transfer protein, putative s...	
259948_at1.236	AT1G71350 [MIPS, Tair, Tigr]eukaryotic translation initiation factor SU11 family protein weak similarity to ...	
257880_at1.190	AT3G16910 [MIPS, Tair, Tigr]AMP-dependent synthetase and ligase family protein similar to AMP-binding protei...	
253330_at1.004	AT4G33530 [MIPS, Tair, Tigr]potassium transporter family protein similar to K+ transporter HAK5 [Arabidopsis...	
!! 251457_s_at1.019	AT3G60160 [MIPS, Tair, Tigr][AT3G60160, ABC transporter family protein similar to ATP-binding cassette tran...	
	AT3G60970 [MIPS, Tair, Tigr][AT3G60970, ABC transporter family protein ABC transporter-like proteins]	
258497_at0.949	AT3G02380 [MIPS, Tair, Tigr]zinc finger protein CONSTANS-LIKE 2 (COL2) identical to putative flowering-time ...	
251885_at0.935	AT3G54050 [MIPS, Tair, Tigr]fructose-1,6-bisphosphatase, putative / D-fructose-1,6-bisphosphate 1-phosphohyd...	
254535_at0.778	AT4G19710 [MIPS, Tair, Tigr]bifunctional aspartate kinase/homoserine dehydrogenase, putative / AK-HSDH, puta...	
262407_at1.966	AT1G34630 [MIPS, Tair, Tigr]expressed protein]	
Hormone: ABA (-)		

Nicht auf dem Chip:

145

## 7.8 Gene der ABF1\_O2\_201-250bp Prädiktionsgruppe (ohne redundante Gene, die in mehreren ABF1-O2-Gruppen auftreten)

Treatment			
Hormone: ABA (+)			
Probeset		AGI/Links/Annotation	
256786_at	0.786	AT3G13740	[MIPS, Tair, Tigr]URF 4-related similar to URF 4 (GI:49224) (Swiss-Prot:Q05370)[Synechococcus sp.]...
251264_at	0.843	AT3G62120	[MIPS, Tair, Tigr]tRNA synthetase class II (G, H, P and S) family protein similar to SP P07814 Bif...
264839_at	0.952	AT1G03630	[MIPS, Tair, Tigr]protochlorophyllide reductase C, chloroplast / PCR C / NADPH-protochlorophyllide...
246454_at	0.952	AT5G16710	[MIPS, Tair, Tigr]dehydroascorbate reductase, putative Strong similarity to dehydroascorbate reduc...
264039_at	0.577	AT2G03740	[MIPS, Tair, Tigr]late embryogenesis abundant domain-containing protein / LEA domain-containing pr...
252736_at	1.160	AT3G43210	[MIPS, Tair, Tigr]kinesin motor family protein (NACK2) contains Pfam profile: PF00225 kinesin moto...
251941_at	1.063	AT3G53470	[MIPS, Tair, Tigr]expressed protein ribosomal protein S25, cytosolic, Arabidopsis thaliana, PIR:T0...
264903_at	1.265	AT1G23190	[MIPS, Tair, Tigr]phosphoglucomutase, cytoplasmic, putative / glucose phosphomutase, putative stro...
246595_at	1.285	AT5G14780	[MIPS, Tair, Tigr]formate dehydrogenase (FDH) identical to GI:7677266]
250978_at	2.912	AT5G03080	[MIPS, Tair, Tigr]phosphatidic acid phosphatase-related / PAP2-related contains Pfam profile PF015...
! 260727_at	11.676	AT1G48100	[MIPS, Tair, Tigr][AT1G48100, glycoside hydrolase family 28 protein / polygalacturonase (pectinas...
		AT1G48110	[MIPS, Tair, Tigr] contains PF00295: Glycosyl hydrolases family 28 (polygalacturonases)]
Hormone: ABA (-)			
		Control	

5 ABA-induzierte Gene  
 3 ABA-reprimierte Gene  
 3 nicht regulierte Gene

## 7.9 Kontrollgruppen mit angepasster Gruppengröße für Gene mit ABF1-TBP-Colokalisation aber ohne klassisches G-Box ABRE

ABA\_Random\_27\_1:

Treatment		
Hormone: ABA (+)		
Probeset	AGI/Links/Annotation	
256278_at1.028	AT3G12130 [MIPS, Tair, Tigr]KH domain-containing protein / zinc finger (CCH type) family protein]	
247196_at1.021	AT5G65510 [MIPS, Tair, Tigr]ovule development protein, putative similar to AINTEGUMENTA (GI:1209099) [Arabid...	
266504_at1.009	AT2G47820 [MIPS, Tair, Tigr]expressed protein]	
253146_at1.014	AT4G35590 [MIPS, Tair, Tigr]RWP-RK domain-containing protein low similarity to minus dominance protein [Chla...	
254771_at1.093	AT4G13380 [MIPS, Tair, Tigr]heavy-metal-associated domain-containing protein low similarity to wound-respons...	
250324_at1.083	AT5G12040 [MIPS, Tair, Tigr]carbon-nitrogen hydrolase family protein similar to Nit protein 2 [Homo sapiens]...	
267600_at0.912	AT2G33040 [MIPS, Tair, Tigr]ATP synthase gamma chain, mitochondrial (ATPC) identical to SPIQ96250 ATP syntha...	
259244_at0.895	AT3G07650 [MIPS, Tair, Tigr]zinc finger (B-box type) family protein similar to zinc finger protein GB:BAA332...	
262173_at0.930	AT1G74920 [MIPS, Tair, Tigr]betaine-aldehyde dehydrogenase, putative identical to betaine-aldehyde dehydroge...	
257715_at1.237	AT3G12750 [MIPS, Tair, Tigr]zinc transporter (ZIP1) identical to putative zinc transporter GB:AAC24197 from ...	
246158_at1.248	AT5G19855 [MIPS, Tair, Tigr]expressed protein]	
267278_at1.182	AT2G19350 [MIPS, Tair, Tigr]expressed protein]	
266591_at0.744	AT2G46225 [MIPS, Tair, Tigr]expressed protein]	
258842_at0.751	AT3G04680 [MIPS, Tair, Tigr]pre-mRNA cleavage complex family protein similar to Pre-mRNA cleavage complex II...	
256583_at0.756	AT3G28850 [MIPS, Tair, Tigr]glutaredoxin family protein]	
246321_at0.760	AT1G16640 [MIPS, Tair, Tigr]transcriptional factor B3 family protein low similarity to reproductive meristem...	
!! 261555_s_at0.724	AT1G63230 [MIPS, Tair, Tigr][AT1G63230, pentatricopeptide (PPR) repeat-containing protein contains Pfam pr...	
258884_at0.723	AT1G63630 [MIPS, Tair, Tigr][AT1G63630, pentatricopeptide (PPR) repeat-containing protein contains Pfam pro...	
251026_at0.721	AT3G10050 [MIPS, Tair, Tigr]threonine ammonia-lyase / threonine dehydratase / threonine deaminase (OMR1) id...	
251592_at0.684	AT5G02200 [MIPS, Tair, Tigr]phytochrome A specific signal transduction component-related contains weak simil...	
251404_at0.786	AT3G57670 [MIPS, Tair, Tigr]zinc finger (C2H2 type) protein (WIP2) identical to WIP2 protein [Arabidopsis th...	
245750_at0.816	AT3G60310 [MIPS, Tair, Tigr]expressed protein]	
247732_at0.605	AT1G51060 [MIPS, Tair, Tigr]histone H2A, putative similar to histone H2A GI:7595337 from Arabidopsis thalian...	
255345_at1.373	AT5G59600 [MIPS, Tair, Tigr]pentatricopeptide (PPR) repeat-containing protein contains INTERPRO:IPR002885 PP...	
253610_at1.417	AT4G04460 [MIPS, Tair, Tigr]aspartyl protease family protein contains Pfam profiles: PF00026 eukaryotic aspa...	
250873_at1.550	AT4G30200 [MIPS, Tair, Tigr]expressed protein contains weak similarities to Pfam profiles: PF00041 Fibronect...	
!! 249616_s_at2.487	AT5G03980 [MIPS, Tair, Tigr]GDSL-motif lipase/hydrolase family protein similar to lipase [Arabidopsis thalia...	
	AT5G37440 [MIPS, Tair, Tigr][AT5G37440, DNAJ heat shock N-terminal domain-containing protein contains Pfam...	
	AT5G37750 [MIPS, Tair, Tigr][AT5G37750, DNAJ heat shock N-terminal domain-containing protein contains Pfam ...	
Hormone: ABA (-)		
		Control

7 Gene ABA-induziert  
11 Gene ABA-reprimiert  
9 Gene nicht reguliert

## ABA\_Random\_27\_2:

Treatment		Hormone: ABA (+)			
Probeset		AGI/Links/Annotation			
! 266679_at	1.299	AT2G03270	[MIPS, Tair, Tigr]	[[AT2G03270, DNA-binding protein, putative similar to Swiss-Prot:Q60560 DNA-bin...	
258912_at	1.166	AT2G03280	[MIPS, Tair, Tigr]	[AT2G03280, expressed protein contains Pfam PF03138: Plant protein family. The ...	
264837_at	0.975	AT3G06460	[MIPS, Tair, Tigr]	GNS1/SUR4 membrane family protein similar to SP P25358 Elongation of fatty acids...	
253283_at	0.969	AT1G03600	[MIPS, Tair, Tigr]	photosystem II family protein similar to SP:P74367 {Synechocystis sp.}; similar ...	
259914_at	0.962	AT4G34090	[MIPS, Tair, Tigr]	expressed protein]	
264439_at	0.983	AT1G72640	[MIPS, Tair, Tigr]	expressed protein]	
261818_at	0.983	AT1G27450	[MIPS, Tair, Tigr]	adenine phosphoribosyltransferase 1 (APT1) nearly identical to SP P31166 Adenine...	
245480_at	0.981	AT1G11390	[MIPS, Tair, Tigr]	ABC1 family protein contains Pfam domain, PF03109: ABC1 family]	
263338_at	1.014	AT4G16150	[MIPS, Tair, Tigr]	calmodulin-binding protein similar to anther ethylene-upregulated calmodulin-bin...	
261517_at	0.895	AT2G05000	[MIPS, Tair, Tigr]	hypothetical protein and genefinder]	
257707_at	0.928	AT1G71780	[MIPS, Tair, Tigr]	expressed protein]	
250628_at	1.044	AT3G12650	[MIPS, Tair, Tigr]	expressed protein]	
246056_at	1.072	AT5G07380	[MIPS, Tair, Tigr]	hypothetical protein]	
267519_at	0.738	AT5G08390	[MIPS, Tair, Tigr]	transducin family protein / WD-40 repeat family protein similar to katanin p80 s...	
251285_at	0.720	AT2G30470	[MIPS, Tair, Tigr]	transcriptional factor B3 family protein low similarity to FUSCA3 [Arabidopsis t...	
253784_at	0.765	AT3G61770	[MIPS, Tair, Tigr]	expressed protein contains Pfam domain, PF02681: Uncharacterized BCR, COG1963]	
261346_at	0.685	AT4G28610	[MIPS, Tair, Tigr]	myb family transcription factor, putative / phosphate starvation response regula...	
258155_at	0.702	AT1G79720	[MIPS, Tair, Tigr]	aspartyl protease family protein contains Pfam domain, PF00026: eukaryotic aspar...	
254763_at	0.660	AT3G18130	[MIPS, Tair, Tigr]	guanine nucleotide-binding family protein / activated protein kinase C receptor ...	
248159_at	0.813	AT4G13170	[MIPS, Tair, Tigr]	60S ribosomal protein L13A (RPL13aC) ribosomal protein L13a -Lupinus luteus,PID:...	
247303_at	0.852	AT5G54460	[MIPS, Tair, Tigr]	wound-responsive protein-related contains weak similarity to KED [Nicotiana taba...	
252547_at	0.525	AT5G63890	[MIPS, Tair, Tigr]	histidinol dehydrogenase, putative / HDH, putative strong similarity to SPIP2422...	
247314_at	0.552	AT3G45840	[MIPS, Tair, Tigr]	DC1 domain-containing protein contains Pfam profile PF03107: DC1 domain]	
259511_at	0.497	AT5G64000	[MIPS, Tair, Tigr]	3'(2'),5'-bisphosphate nucleotidase, putative / inositol polyphosphate 1-phospha...	
263146_at	0.276	AT1G12520	[MIPS, Tair, Tigr]	superoxide dismutase copper chaperone, putative similar to copper chaperone for ...	
263545_at	1.931	AT1G53940	[MIPS, Tair, Tigr]	GDSL-motif lipase/hydrolase family protein similar to lipase GI:1145627 from (Ar...	
!! 257400_s_at	1.780	AT2G21560	[MIPS, Tair, Tigr]	expressed protein contains weak similarity to reticulocyte-binding protein 2 hom...	
		AT1G23810	[MIPS, Tair, Tigr]	[AT1G23810, paired amphipathic helix repeat-containing protein low similarity t...	
		AT1G24250	[MIPS, Tair, Tigr]	[AT1G24250, paired amphipathic helix repeat-containing protein weak similarity...	
		Hormone: ABA (-)			
Control					

4 Gene ABA-induziert  
 12 Gene ABA-reprimiert  
 11 Gene nicht reguliert

## ABA\_Random\_27\_3:

Treatment			
Hormone: ABA (+)			
Probeset		AGI/Links/Annotation	
257985_a	1.051	AT3G20810	[MIPS, Tair, Tigr]transcription factor jumonji (jmiC) domain-containing protein contains Pfam doma...
255538_a	1.031	AT4G01680	[MIPS, Tair, Tigr]myb family transcription factor (MYB55)
253748_a	0.994	AT4G28990	[MIPS, Tair, Tigr]RNA-binding protein-related contains weak similarity to Swiss-Prot:Q01844 RNA-bi...
246115_a	1.001	AT5G20300	[MIPS, Tair, Tigr]chloroplast outer membrane protein, putative similar to chloroplast protein impo...
257799_a	1.169	AT3G15890	[MIPS, Tair, Tigr]protein kinase family protein contains protein kinase domain, Pfam:PF00069]
247836_a	1.155	AT5G57860	[MIPS, Tair, Tigr]ubiquitin family protein contains Pfam profile PF00240: Ubiquitin family]
255756_a	0.782	AT1G19940	[MIPS, Tair, Tigr]glycosyl hydrolase family 9 protein similar to endo-beta-1,4-D-glucanase Gl:4165...
253909_a	0.767	AT4G27270	[MIPS, Tair, Tigr]quinone reductase family protein similar to 1,4-benzoquinone reductase [Phaneroc...
262451_a	0.752	AT1G11130	[MIPS, Tair, Tigr]leucine-rich repeat family protein / protein kinase family protein contains leuc...
245053_a	0.696	AT2G26450	[MIPS, Tair, Tigr]pectinesterase family protein contains Pfam profiles: PF01095 pectinesterase,PF0...
250786_a	0.883	AT5G05540	[MIPS, Tair, Tigr]exonuclease family protein contains exonuclease domain, Pfam:PF00929]
250112_a	0.901	AT5G16300	[MIPS, Tair, Tigr]expressed protein]
253141_a	0.920	AT4G35440	[MIPS, Tair, Tigr]voltage-gated chloride channel family protein contains Pfam profile PF00654: Vol...
245574_a	0.933	AT4G14750	[MIPS, Tair, Tigr]calmodulin-binding family protein contains Pfam profile PF00612: IQ calmodulin-b...
264394_a	0.844	AT1G11860	[MIPS, Tair, Tigr]aminomethyltransferase, putative similar to aminomethyltransferase, mitochondria...
245548_a	0.846	AT4G15310	[MIPS, Tair, Tigr]cytochrome P450-related contains weak similarity to Pfam profile: PF00067: Cytoc...
250111_a	0.849	AT5G16290	[MIPS, Tair, Tigr]acetolactate synthase small subunit, putative similar to gi:5931761 from Nicotia...
253467_a	0.817	AT4G32140	[MIPS, Tair, Tigr]expressed protein]
257349_a	0.501	AT2G30630	[MIPS, Tair, Tigr]expressed protein similar to H1-1flk [Arabidopsis thaliana] Gl:388260]
250443_a	0.498	AT5G10520	[MIPS, Tair, Tigr]protein kinase family protein contains protein kinase domain, INTERPRO:IPR000719...
261380_a	0.518	AT1G05400	[MIPS, Tair, Tigr]hypothetical protein contains Pfam profile PF04396: Protein of unknown function,...
267160_a	1.493	AT2G37670	[MIPS, Tair, Tigr]WD-40 repeat family protein contains 6 WD-40 repeats (PF00400); similar to rab1...
259592_a	1.547	AT1G27950	[MIPS, Tair, Tigr]lipid transfer protein-related low similarity to lipid transfer protein Picea ab...
254759_a	1.399	AT4G13180	[MIPS, Tair, Tigr]short-chain dehydrogenase/reductase (SDR) family protein similar to short-chain ...
262740_a	1.812	AT1G28590	[MIPS, Tair, Tigr]lipase, putative similar to lipase [Arabidopsis thaliana] Gl:1145627; contains l...
249707_a	1.813	AT5G35640	[MIPS, Tair, Tigr]hypothetical protein contains Pfam profile PF04396: Protein of unknown function,...
262803_a	2.499	AT1G21000	[MIPS, Tair, Tigr]zinc-binding family protein similar to zinc-binding protein [Pisum sativum] Gl:1...
Hormone: ABA (-)			
		Control	

8 Gene ABA-induziert  
 11 Gene ABA-reprimiert  
 8 Gene nicht reguliert

## ABA\_Random\_27\_4:

Treatment		
Hormone: ABA (+)		
Probeset	AGI/Links/Annotation	
250172_at	0.863	AT5G14330 [MIPS, Tair, Tigr]expressed protein]
245718_at	0.848	AT5G04110 [MIPS, Tair, Tigr]DNA topoisomerase II family protein siimilar to DNA topoisomerase II subunit B [...
258821_at	0.822	AT3G07230 [MIPS, Tair, Tigr]wound-responsive protein-related similar to wound-induced basic protein SP:Q0902...
253988_at	0.897	AT4G26310 [MIPS, Tair, Tigr]elongation factor P (EF-P) family protein similar to SP Q45288 Elongation factor...
!! 253340_s	at0.901	AT4G33260 [MIPS, Tair, Tigr][[AT4G33260, WD-40 repeat family protein contains 6 WD-40 repeats (PF00400); WD...
		AT4G33270 [MIPS, Tair, Tigr][AT4G33270, WD-40 repeat family protein contains 6 WD-40 repeats (PF00400); WD-...
265430_at	0.880	AT2G20700 [MIPS, Tair, Tigr]expressed protein]
258502_at	0.753	AT3G02490 [MIPS, Tair, Tigr]pentatricopeptide (PPR) repeat-containing protein contains Pfam profile PF01535:...
253743_at	0.746	AT4G28940 [MIPS, Tair, Tigr]nucleosidase-related contains weak similarity to MTA/SAH nucleosidase (P46). (Sw...
253854_at	0.738	AT4G27900 [MIPS, Tair, Tigr]expressed protein]
264286_at	0.727	AT1G61870 [MIPS, Tair, Tigr]pentatricopeptide (PPR) repeat-containing protein contains Pfam profile PF01535:...
253992_at	0.773	AT4G26060 [MIPS, Tair, Tigr]expressed protein]
245960_at	0.773	AT5G19660 [MIPS, Tair, Tigr]subtilase family protein contains Pfam profile: PF00082 subtilase family]
259189_at	0.579	AT3G01700 [MIPS, Tair, Tigr]arabinogalactan-protein (AGP11) identical to gi 10880499 gb AAG24279]
252650_at	0.597	AT3G44690 [MIPS, Tair, Tigr]expressed protein]
261007_at	1.200	AT1G26400 [MIPS, Tair, Tigr]FAD-binding domain-containing protein similar to SP P30986 reticuline oxidase pr...
251742_at	1.131	AT3G56050 [MIPS, Tair, Tigr]protein kinase family protein contains Pfam domain, PF00069: Protein kinase doma...
251929_at	1.027	AT3G53920 [MIPS, Tair, Tigr]RNA polymerase sigma subunit SigC (sigC) / sigma factor 3 (SIG3) identical to si...
249564_at	1.025	AT5G38400 [MIPS, Tair, Tigr]hypothetical protein]
266724_at	1.046	AT2G03180 [MIPS, Tair, Tigr]hypothetical protein]
!! 254471_s	at0.967	AT4G20500 [MIPS, Tair, Tigr][ similar to YEAST NUF1 protein (Spindle poly body spacer protein SPC110) (SP:P3...
		AT4G20730 [MIPS, Tair, Tigr][AT4G20730, filament protein-related similar to Cytadherence high molecular wei...
266270_at	1.300	AT2G29470 [MIPS, Tair, Tigr]glutathione S-transferase, putative similar to glutathione S-transferase [Euphor...
261236_at	1.379	AT1G32880 [MIPS, Tair, Tigr]importin alpha-1 subunit, putative similar to importin alpha-1 subunit (Karyophe...
263139_at	1.499	AT1G65110 [MIPS, Tair, Tigr]ubiquitin carboxyl-terminal hydrolase-related contains Pfam profiles PF00443: Ub...
245204_at	1.477	AT5G12270 [MIPS, Tair, Tigr]oxidoreductase, 2OG-Fe(II) oxygenase family protein similarity to ripening prote...
259986_at	1.687	AT1G75050 [MIPS, Tair, Tigr]thaumatin-like protein, putative / pathogenesis-related protein, putative simila...
245154_at	3.094	AT5G12460 [MIPS, Tair, Tigr]fringe-related protein similarity to predicted proteins + similar to hypothetica...
253263_at	7.752	AT4G34000 [MIPS, Tair, Tigr]ABA-responsive element-binding protein / abscisic acid responsive elements-bindi...
Hormone: ABA (-)		
Control		

8 Gene ABA-induziert  
 11 Gene ABA-reprimiert  
 8 Gene nicht reguliert

## ABA\_Random\_27\_5:

Treatment			
Hormone: ABA (+)			
Probeset	AGI/Links/Annotation		
263170_at0.831	AT1G03000 [MIPS, Tair, Tigr]AAA-type ATPase family protein contains Pfam domain, PF00004: ATPase, AAA family...		
251305_at0.830	AT3G62030 [MIPS, Tair, Tigr]peptidyl-prolyl cis-trans isomerase, chloroplast / cyclophilin / rotamase / cycl...		
263877_at0.779	AT2G21780 [MIPS, Tair, Tigr]expressed protein]		
262441_at0.897	AT1G47720 [MIPS, Tair, Tigr]expressed protein contains Pfam PF05329: Protein of unknown function (DUF731)]		
258471_at0.899	AT3G06030 [MIPS, Tair, Tigr]NPK1-related protein kinase, putative (ANP3) similar to protein kinase [Nicotian...		
258495_at0.903	AT3G02690 [MIPS, Tair, Tigr]integral membrane family protein similar to PecM protein (GI:5852331) (Vogesella...		
250568_at0.882	AT5G08110 [MIPS, Tair, Tigr]DEAD/DEAH box helicase, putative several putative ATP-dependent helicases]		
265367_at0.720	AT2G13270 [MIPS, Tair, Tigr]hypothetical protein]		
246776_at0.725	AT5G27550 [MIPS, Tair, Tigr]kinesin motor protein-related kinesin-like heavy chain - Arabidopsis thaliana, E...		
257197_at0.661	AT3G23800 [MIPS, Tair, Tigr]selenium-binding family protein contains Pfam profile: PF05694 56kDa selenium bi...		
251677_at0.552	AT3G56980 [MIPS, Tair, Tigr]basic helix-loop-helix (bHLH) family protein]		
! 248851_s_at0.473	AT5G46490 [MIPS, Tair, Tigr]disease resistance protein (TIR-NBS class), putative domain signature TIR-NBS ex...		
264932_at1.178	AT1G61240 [MIPS, Tair, Tigr]expressed protein contains Pfam profile PF05212: Protein of unknown function (DU...		
!! 245285_s_at1.114	AT4G14030 [MIPS, Tair, Tigr][AT4G14030, selenium-binding protein, putative contains Pfam profile PF05694: 5...		
	AT4G14040 [MIPS, Tair, Tigr][ similar to selenium binding protein (GI:15485232) [Arabidopsis thaliana]		
249950_at1.245	AT5G18910 [MIPS, Tair, Tigr]protein kinase family protein contains protein kinase domain, Pfam:PF00069]		
245547_at1.246	AT4G15300 [MIPS, Tair, Tigr]cytochrome P450 family protein similar to Cytochrome P450 90C1 (ROTUNDIFOLIA3) (...)		
263785_at1.263	AT2G46390 [MIPS, Tair, Tigr]expressed protein]		
265459_at0.975	AT2G46540 [MIPS, Tair, Tigr]expressed protein]		
!! 264067_x_at1.003	AT2G28010 [MIPS, Tair, Tigr][AT2G28010, aspartyl protease family protein contains Pfam domain, PF00026: eu...		
	AT2G28030 [MIPS, Tair, Tigr][AT2G28030, aspartyl protease family protein contains Pfam domain, PF00026: euk...		
258041_at1.039	AT3G21175 [MIPS, Tair, Tigr]zinc finger (GATA type) family protein similar to zinc finger protein ZIM gi:891...		
249068_at1.050	AT5G43980 [MIPS, Tair, Tigr]receptor-like protein kinase-related similar to receptor-like protein kinase hom...		
250395_at1.062	AT5G10950 [MIPS, Tair, Tigr]cylicin-related low similarity to SP Q28092 Cylicin II (Multiple-band polypeptid...		
264972_at1.657	AT1G67370 [MIPS, Tair, Tigr]meiotic asynaptic mutant 1 (ASY1) identical to meiotic asynaptic mutant 1 [Arabi...		
258169_at1.461	AT3G21590 [MIPS, Tair, Tigr]senescence/dehydration-associated protein-related similar to senescence-associat...		
265769_at2.283	AT2G48090 [MIPS, Tair, Tigr]expressed protein weak similarity to protein kinase]		
261899_at4.754	AT1G80820 [MIPS, Tair, Tigr]cinnamoyl-CoA reductase, putative identical to CCR2 (GI:12407990), similar to ci...		
252321_at10.557	AT3G48510 [MIPS, Tair, Tigr]expressed protein]		
Hormone: ABA (-)			
		Control	

9 Gene ABA-induziert  
8 Gene ABA-reprimiert  
10 Gene nicht reguliert

### 7.9.1 t-Test auf Signifikanz nach Subtraktion der Gene mit klassischem ABRE

t-Test auf Signifikanz:

Induzierte Gene in der Prädiktionsgruppe nach Abzug der vier induzierten Gene mit klassischem ABRE: 17 Gene

Durchschnittliche Anzahl ABA-induzierter Gene in den Kontrollgruppen: 7,2  
Standardabweichung: 1,92

$$t = \frac{17 - 7,2}{\frac{1,92}{\sqrt{5}}} = 11,41$$

Grenzwert bei n=5 und  $\alpha = 0,005$ : 5,598



## 7.10 ABF1\_TSS\_0\_50bp-Gruppe und Vergleichsgruppen zufällig ausgewählter Gene

Folgende 29 Gene wurden in der ABF1-TSS-Colokalisationsanalyse identifiziert:

At1g03030  
At1g06110  
At1g09870  
At1g26690  
At1g56330  
At1g60620  
At1g61065  
At1g78390  
At2g04100  
At2g21860  
At2g30200  
At2g42790  
At2g46270  
At3g03160  
At3g08730  
At3g16140  
At3g21055  
At3g45600  
At3g48510  
At3g50820  
At4g03200  
At4g04870  
At4g19710  
At4g34870  
At5g10070  
At5g16390  
At5g38590  
At5g42740  
At5g55070

Nach Subtraktion aller Gene mit klassischem ABRE oder ABF1-TBP-Colokalisation blieben die folgenden 11 Gene die auf ihre ABA-Induzierbarkeit untersucht wurden:

Treatment			
Hormone: ABA (+)			
Probeset		AGI/Links/Annotation	
267308_at	0.786	AT2G30200	[MIPS, Tair, Tigr]expressed protein]
264924_at	0.804	AT1G60620	[MIPS, Tair, Tigr]DNA-directed RNA polymerase, putative identical to RNA polymerase subunit [Arabi...
256979_at	0.745	AT3G21055	[MIPS, Tair, Tigr]photosystem II 5 kD protein, putative identical to Swiss-Prot:Q39195 photosystem...
263873_at	0.866	AT2G21860	[MIPS, Tair, Tigr]violaxanthin de-epoxidase-related contains weak similarity to violaxanthin de-ep...
259725_at	1.038	AT1G61065	[MIPS, Tair, Tigr]expressed protein]
249163_at	1.077	AT5G42740	[MIPS, Tair, Tigr]glucose-6-phosphate isomerase, cytosolic (PGIC) identical to SP P34795 Glucose-6...
250125_at	1.137	AT5G16390	[MIPS, Tair, Tigr]biotin carboxyl carrier protein 1 (BCCP1) identical to biotin carboxyl carrier p...
248088_at	1.134	AT5G55070	[MIPS, Tair, Tigr]2-oxoacid dehydrogenase family protein similar to SP Q01205 Dihydrolipoamide suc...
! 258677_at	1.127	AT3G08720	[MIPS, Tair, Tigr][AT3G08720, serine/threonine protein kinase (PK19) identical to serine/threonin...
		AT3G08730	[MIPS, Tair, Tigr][AT3G08730, serine/threonine protein kinase (PK1) (PK6) identical to serine/th...
260949_at	1.329	AT1G06110	[MIPS, Tair, Tigr]F-box family protein contains similarity to F-box protein FBX3 GI:6103643 from [...
252591_at	1.504	AT3G45600	[MIPS, Tair, Tigr]senescence-associated family protein similar to senescence-associated protein 5 ...
Hormone: ABA (-)			
		Control	

2 Gene ABA-induziert  
 4 Gene ABA-reprimiert  
 5 Gene nicht reguliert

Random11\_1:

Treatment			
Hormone: ABA (+)			
Probeset	AGI/Links/Annotation		
265676_at0.872	AT2G32070 [MIPS, Tair, Tigr]CCR4-NOT transcription complex protein, putative similar to SWISS-PROT:Q60809 CC...		
249427_at0.859	AT5G39850 [MIPS, Tair, Tigr]40S ribosomal protein S9 (RPS9C) 40S ribosomal protein S9 - Chlamydomonas sp.,EM...		
! 248045_at0.841	AT5G56000 [MIPS, Tair, Tigr][AT5G56000, heat shock protein 81-4 (HSP81-4) nearly identical to heat shock p...		
	AT5G56010 [MIPS, Tair, Tigr][AT5G56010, heat shock protein, putative strong similarity to SP P55737 Heat s...		
	AT5G56030 [MIPS, Tair, Tigr][AT5G56030, heat shock protein 81-2 (HSP81-2) nearly identical to SP P55737 Hea...		
255504_at0.937	AT4G02200 [MIPS, Tair, Tigr]drought-responsive family protein similar to drought-induced mRNA, Di19 [Arabido...		
252502_at1.016	AT3G46900 [MIPS, Tair, Tigr]copper transporter, putative similar to SP Q39065 Copper transporter 1 (COPT1) {...		
245430_at1.015	AT4G17060 [MIPS, Tair, Tigr]expressed protein]		
249864_at1.071	AT5G22830 [MIPS, Tair, Tigr]magnesium transporter CorA-like family protein weak similarity to SP Q01926 RNA ...		
256602_at0.459	AT3G28310 [MIPS, Tair, Tigr]expressed protein similar to At14a (TIGR_Ath1:At3g28290, TIGR_Ath1:At3g28300) [A...		
262356_at1.574	AT1G73000 [MIPS, Tair, Tigr]hypothetical protein]		
!! 256318_s_at1.695	AT1G35940 [MIPS, Tair, Tigr][AT1G35940, AT hook motif-containing protein-related very low similarity to SP...		
	AT5G34960 [MIPS, Tair, Tigr][AT5G34960, hypothetical protein includes At5g34960, At2g14450, At1g35920]		
263829_at3.573	AT2G40435 [MIPS, Tair, Tigr]expressed protein]		
Hormone: ABA (-)			
		Control	

3 Gene ABA-induziert  
3 Gene ABA-reprimiert  
5 Gene nicht reguliert

Random11\_2:

Treatment			
Hormone: ABA (+)			
Probeset		AGI/Links/Annotation	
257818_a	0.659	AT3G25120	[MIPS, Tair, Tigr]mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family...
247796_a	0.697	AT5G58782	[MIPS, Tair, Tigr]dehydrodolichyl diphosphate synthase, putative / DEDOL-PP synthase, putative sim...
259021_a	0.835	AT3G07540	[MIPS, Tair, Tigr]formin homology 2 domain-containing protein / FH2 domain-containing protein cont...
263050_a	1.055	AT2G05170	[MIPS, Tair, Tigr]vacuolar protein sorting 11 family protein / VPS11 family protein similar to Vac...
253574_a	1.077	AT4G31030	[MIPS, Tair, Tigr]expressed protein]
262046_a	1.015	AT1G79960	[MIPS, Tair, Tigr]ovate protein-related contains TIGRFAM TIGR01568 : uncharacterized plant-specifi...
248722_a	1.021	AT5G47810	[MIPS, Tair, Tigr]phosphofructokinase family protein similar to phosphofructokinase [Amycolatopsis...
262346_a	0.948	AT1G63980	[MIPS, Tair, Tigr]D111/G-patch domain-containing protein contains Pfam profile PF01585: G-patch do...
250854_a	0.946	AT5G04710	[MIPS, Tair, Tigr]aspartyl aminopeptidase, putative similar to SP Q9ULA0 Aspartyl aminopeptidase (...)
267134_a	1.397	AT2G23450	[MIPS, Tair, Tigr]protein kinase family protein contains protein kinase domain, Pfam:PF00069]
245782_a	1.481	AT3G09630	[MIPS, Tair, Tigr]60S ribosomal protein L4/L1 (RPL4A) strong similarity to 60S ribosomal protein L...
Hormone: ABA (-)			
Control			

- 2 Gene ABA-induziert
- 3 Gene ABA-reprimiert
- 6 Gene nicht reguliert

Probeset	Hormone: ABA (+)	Hormone: ABA (-)	AGI/Links/Annotation
261689_a	0.874		AT1G50140 [MIPS, Tair, Tigr]AAA-type ATPase family protein contains Pfam domain, PF00004: ATPase, AAA family...
260250_a	0.920		AT1G74260 [MIPS, Tair, Tigr]AIR synthase-related family protein contains Pfam profiles: PF00586 AIR synthase...
266523_a	0.809		AT2G16950 [MIPS, Tair, Tigr]importin beta-2 subunit family protein similar to SP Q92973 Importin beta-2 subu...
264171_a	0.819		AT1G02100 [MIPS, Tair, Tigr]leucine carboxyl methyltransferase family protein contains Pfam PF04072: Leucine...
250064_a	0.795		AT5G17890 [MIPS, Tair, Tigr]LIM domain-containing protein / disease resistance protein-related low similarit...
266109_a	0.704		AT2G37890 [MIPS, Tair, Tigr]mitochondrial substrate carrier family protein contains Pfam profile: PF00153 mi...
248704_a	0.708		AT5G48450 [MIPS, Tair, Tigr]multi-copper oxidase type I family protein contains Pfam profile: PF00394 Multic...
257881_a	1.129		AT3G17180 [MIPS, Tair, Tigr]serine carboxypeptidase S10 family protein similar to serine carboxypeptidase II...
248034_a	1.118		AT5G55910 [MIPS, Tair, Tigr]protein kinase, putative contains protein kinase domain, Pfam:PF00069]
267459_a	1.103		AT2G33850 [MIPS, Tair, Tigr]expressed protein contains 1 transmembrane domain; similar to Protein E6 (Swiss-...
261102_a	1.048		AT1G62880 [MIPS, Tair, Tigr]cornichon family protein contains Pfam profile: PF03311 cornichon protein]

157

Random11\_4:

Treatment	
Probeset	AGI/Links/Annotation
263353_at1.040	AT2G22070 [MIPS, Tair, Tigr]pentatricopeptide (PPR) repeat-containing protein contains INTERPRO:IPR002885 PP...
258150_at1.043	AT3G18160 [MIPS, Tair, Tigr]peroxin-3 family protein contains Pfam domain, PF04882: Peroxin-3]
249068_at1.050	AT5G43980 [MIPS, Tair, Tigr]receptor-like protein kinase-related similar to receptor-like protein kinase hom...
264356_at1.066	AT1G03190 [MIPS, Tair, Tigr]DNA repair protein / transcription factor protein (UVH6) identical to DNA repair...
252058_at1.059	AT3G52470 [MIPS, Tair, Tigr]harpin-induced family protein / HIN1 family protein / harpin-responsive family p...
248829_at0.997	AT5G47130 [MIPS, Tair, Tigr]Bax inhibitor-1 family / BI-1 family similar to SPIQ9LD45 Bax inhibitor-1 (BI-1)...
252187_at1.122	AT3G50850 [MIPS, Tair, Tigr]expressed protein]
246782_at0.918	AT5G27320 [MIPS, Tair, Tigr]expressed protein similar to PrMC3 [Pinus radiata] GI:5487873]
265717_at0.746	AT2G03510 [MIPS, Tair, Tigr]band 7 family protein contains Pfam profile PF01145: SPFH domain / Band 7 family...
263674_at0.682	AT2G04790 [MIPS, Tair, Tigr]expressed protein]
!! 261079_s_at1.919	AT1G07470 [MIPS, Tair, Tigr][AT1G07470, transcription factor IIA large subunit, putative / TFIIA large subu...
	AT1G07480 [MIPS, Tair, Tigr][ contains Pfam profile: PF03153 transcription factor IIA, alpha/beta subunit]
Hormone: ABA (+)	
Hormone: ABA (-)	
Control	

1 ABA-induziertes Gen  
2 ABA-reprimierte Gene  
8 nicht regulierte Gene

Treatment		
Hormone: ABA (+)		
Probeset	AGI/Links/Annotation	
256536_a	0.891 AT1G33330 [MIPS, Tair, Tigr]peptide chain release factor, putative similar to peptide chain release factor 2...	
250493_a	0.919 AT5G09800 [MIPS, Tair, Tigr]U-box domain-containing protein low similarity to immediate-early fungal elicito...	
247689_a	1.000 AT5G59770 [MIPS, Tair, Tigr]expressed protein]	
266731_a	1.081 AT2G03260 [MIPS, Tair, Tigr]EXS family protein / ERD1/XPR1/SYG1 family protein similar to PHO1 protein [Arab...	
250427_a	1.118 AT5G10500 [MIPS, Tair, Tigr]kinase interacting family protein similar to kinase interacting protein 1 (GI:13...	
266007_a	0.659 AT2G37380 [MIPS, Tair, Tigr]expressed protein]	
257137_a	0.719 AT3G28860 [MIPS, Tair, Tigr]multidrug resistance P-glycoprotein, putative similar to mdr-like P-glycoprotein...	
262999_a	0.780 AT1G54330 [MIPS, Tair, Tigr]no apical meristem (NAM) family protein contains Pfam PF02365: No apical meriste...	
258326_a	0.820 AT3G22760 [MIPS, Tair, Tigr]CXC domain containing TSO1-like protein 1 (SOL1) identical to CXC domain contain...	
258066_a	1.576 AT3G25970 [MIPS, Tair, Tigr]pentatricopeptide (PPR) repeat-containing protein contains INTERPRO:IPR002885 PP...	
252872_a	7.174 AT4G40010 [MIPS, Tair, Tigr]serine/threonine protein kinase, putative similar to serine-threonine protein ki...	
Hormone: ABA (-)		

- 159

## Danksagung

Mein besonderer Dank gilt Professor Dr. Reinhard Hehl, meinem Mentor, für die Bereitstellung des Themas dieser Arbeit. Außerdem möchte ich mich bedanken für seine ständige Diskussionsbereitschaft in allen wissenschaftlichen Fragen, die Inspiration zu guten Ideen und seine Fairness gegenüber seinen Doktoranden.

Herrn Professor Wingender danke ich für die Bereitschaft, das Coreferat für diese Arbeit zu übernehmen. Darüber hinaus möchte ich Herrn Wingender für die gute Kooperation zwischen Biobase und der AG Hehl danken. Ohne die TRANSFAC-Ressourcen, die wir großzügiger Weise nutzen durften, wäre diese Arbeit erheblich schwieriger geworden.

Außerdem möchte ich allen Mitarbeitern des Institutes für Genetik für die angenehme Atmosphäre danken. Speziell meiner Kollegin Claudia Galuschka bin ich dankbar für anregende wissenschaftliche Diskussionen und freundschaftliche Gespräche. Meinem Kollegen Alexander Rotthues gilt mein Dank für die gute Kooperation bei der gemeinsamen Betreuung von Studenten.

Zu guter Letzt möchte ich mich bei meiner Familie und meinen Freunden für die Unterstützung bedanken und ihrem Verständnis für „Ausnahmestände“.



## **LEBENS LAUF UND AUSBILDUNGSGANG**

### **Persönliche Daten:**

Name                                Nils Ole Steffens  
Geburtsdatum und -ort:    10.01.1974 in Hamburg  
Staatsangehörigkeit        deutsch  
Familienstand:                Ledig

### **Schulische und berufliche Ausbildung:**

1993	Abitur am Gymnasialzweig der Kooperativen Gesamtschule Bad Bevensen
1993 – 1994	Zivildienst im ökumenischen Gemeindezentrum St. Stephanus in Lüneburg
1994 –2001	Studium der Biologie (Fachrichtung Zellbiologie, Botanik und Genetik) an der Technischen Universität Braunschweig. Abschluß als Diplom-Biologe mit der Note „Sehr gut“ Thema der Diplomarbeit: „Charakterisierung der Wechselwirkung zwischen Raver1 und Vinculin/ Metavinculin“
2001- 2002	Beschäftigung als Doktorand am Fraunhofer-Institut für Toxikologie und Aerosolforschung (ITA), Hannover
2002- 2006	Beschäftigung als Doktorand am Institut für Genetik der TU Braunschweig (AG Professor Hehl)

### **Allgemeine Veröffentlichungen:**

Hüttelmaier S., Harbeck B., Steffens O., Messerschmidt T., Illenberger S., Jockusch  
B.M.: Characterization of the actin binding properties of the vasodilator-stimulated  
phosphoprotein VASP. *FEBS Lett.* (1999) 451(1):68-74.  
Hehl, R., Steffens N.O., Wingender E. (2004) Isolation and analysis of gene  
regulatory sequences. In: Handbook of Plant Biotechnology, Klee H, and Christou P,  
eds., (Wiley and Sons Ltd.), pp. 81-102.